

A0 5905

UNIVERSITE PARIS VII

THESE DE DOCTORAT ES SCIENCES

Présentée par
Jacques NINIO

Sujet de la thèse : ETUDE DE LA STRUCTURE
DE L'ARN DE TRANSFERT PAR DIFFUSION
CENTRALE DES RAYONS X, ET DE SES
IMPLICATIONS BIOLOGIQUES

Soutenue le 17 juin 1971

devant le Jury composé de :

MM.	F. CHAPEVILLE	Président
	P.-G. DE GENNES V. LUZZATI	} Examineurs
	F. GROS A. M. MICHELSON	} Invités



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41556001_0012

A TOUS CEUX QUI M'ONT
ACCORDE LEUR CONFIANCE .

Mathématicien de formation, j'ai été encouragé par Jacques Monod à participer au mouvement de la biologie moléculaire ; je ne regrette pas ce tournant.

J'ai connu ensuite la réalité de la condition de chercheur scientifique au Laboratoire de Génétique Physiologique, puis au Centre de Génétique Moléculaire. Je suis reconnaissant à M. Slonimski de m'avoir accueilli, et à M. V. Luzzati d'avoir dirigé ma thèse avec dévouement. Je suis également reconnaissant à M. Luzzati de m'avoir proposé un sujet de thèse particulièrement intéressant, et riche de possibilités. Qu'il me pardonne si j'ai tenté d'en franchir les limites.

Je remercie tous ceux dans le laboratoire qui ont contribué à maintenir élevée la flamme de la discussion scientifique, et en particulier C. Reiss et P. Saludjian. Que tous les autres également, chercheurs et techniciens soient remerciés pour leur généreuse camaraderie.

Ce travail n'aurait sans doute pas vu le jour sans les dons généreux de M. Yaniv en matériel hautement purifié, les collaborations amicales de A. Favre et P. Claverie. Il a également bénéficié de l'aide pratique du groupe de Mme Grunberg-Manago et de celui de J.-P. Waller, ainsi que d'une initiation à la biochimie conduite par H. Fukuhara et J.-C. Mounolou, qui m'ont consacré une partie précieuse de leur temps.

Je suis particulièrement reconnaissant à Boris Ephrussi et François Gros pour l'ouverture d'esprit qu'ils ont manifesté en m'écoutant, pour leur haute probité intellectuelle et morale.

Enfin je suis reconnaissant aux membres du jury de l'honneur qu'ils me font en acceptant de juger ce travail.

TABLE DES MATIERES.

<u>1-INTRODUCTION.</u>	6
<u>2-TECHNIQUES DE RAYONS X.</u>	14
Notations	
Rappel des équations de la collimation linéaire infinie	
Dispositif expérimental	
Equations de la collimation	
Corrections de collimation	
Calcul des intensités théoriques	
<u>3-MODELES MOLECULAIRES.</u>	31
Généralités sur l'emploi des modèles moléculaires	
Construction des modèles de tARN	
Simulation des modèles moléculaires	
Intensités théoriques correspondant à différents modèles	
<u>4-PROPRIETES DES REPRESENTATIONS D'ACIDES NUCLEIQUES.</u>	43
Introduction	
Propriétés de topologie	
Propriétés d'orientation	
<u>5-PREPARATION DES ECHANTILLONS.</u>	65
<u>6-RESULTATS DES EXPERIENCES DE DIFFUSION CENTRALE DES RAYONS X SUR LES</u>	
<u>ACIDES NUCLEIQUES.</u>	69
ARN ribosomal -eau	
ARN ribosomal et ARN de transfert en glycol	
ADN-eau-magnésium	
tARN non chargé	
Phénoxyacetyl-tARN ^{Val}	
tARN ^{Val} non chargé - tARN ^{Val} irradié	
Détermination du poids moléculaire	
Détermination du rayon de gyration	
Parametres de forme	
Queues de courbes	

<u>7- LE MODELE 4-2-2.</u>	87
<u>8-LA RECONNAISSANCE CODON-ANTICODON.</u>	93
<u>9-CONCLUSION.</u>	100
<u>10-BIBLIOGRAPHIE.</u>	104

- CHAPITRE I -

INTRODUCTION

Pour raconter l'histoire de l'ARN de transfert, il n'est pas nécessaire de consulter des ouvrages poussiéreux à la recherche de sources obscures. Entre la première caractérisation de cette substance, en 1957⁽¹⁾, et la détermination de sa structure cristalline (encore ignorée) il se sera écoulé une quinzaine d'années.

L'idée de l'existence de cette classe de molécules a été mise en circulation par Crick, dès 1955, dans le "club de la cravate" fondé par Gamow. Les raisonnements qui l'ont amené à postuler la présence d'adapteurs dans la traduction du texte génétique sont exposées dans l'article original, dont il existe très peu d'exemplaires. On pourra trouver les premières références ainsi qu'un exposé très vivant des conceptions de l'époque dans un article ultérieur⁽²⁾.

L'argument invoqué par Crick est qu'une reconnaissance stéréochimique directe entre les codons sur l'ADN et les acides aminés est peu plausible, et qu'il faut donc imaginer l'intervention d'adapteurs, et celle d'enzymes d'activation pour fixer l'acide aminé sur l'adapteur.

Les conceptions de Crick se sont trouvées largement confirmées par l'expérience. Mais l'adapteur, au lieu d'être un petit oligonucléotide qui comporterait trois à cinq résidus, est une macromolécule de poids moléculaire 25.000⁽³⁾.

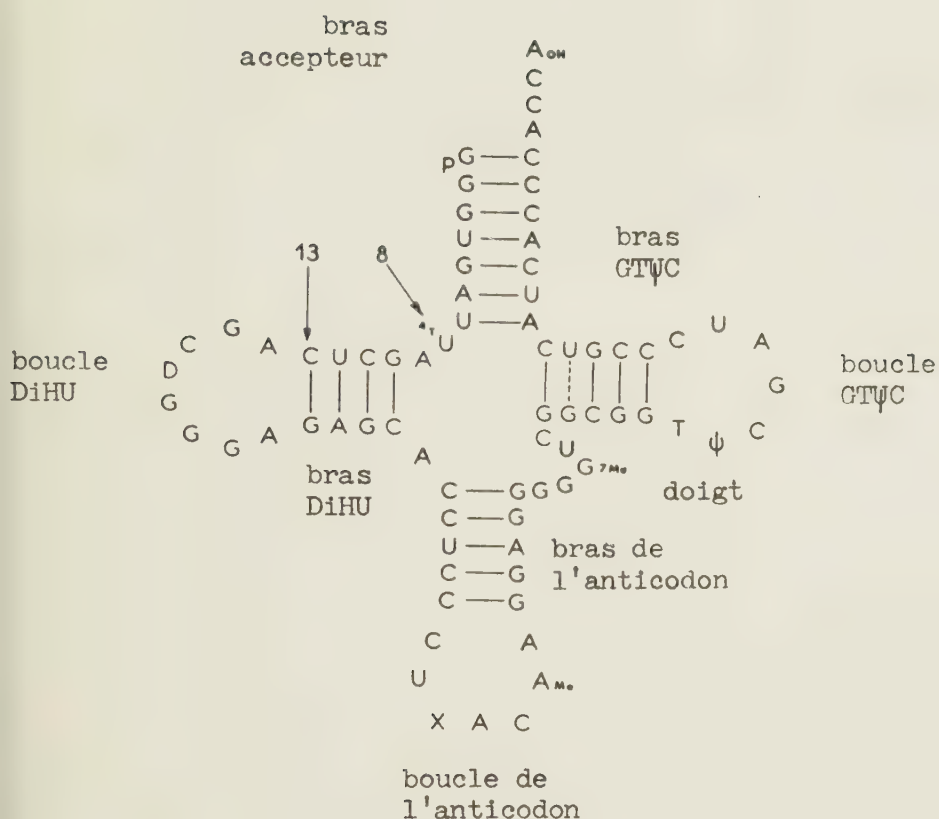
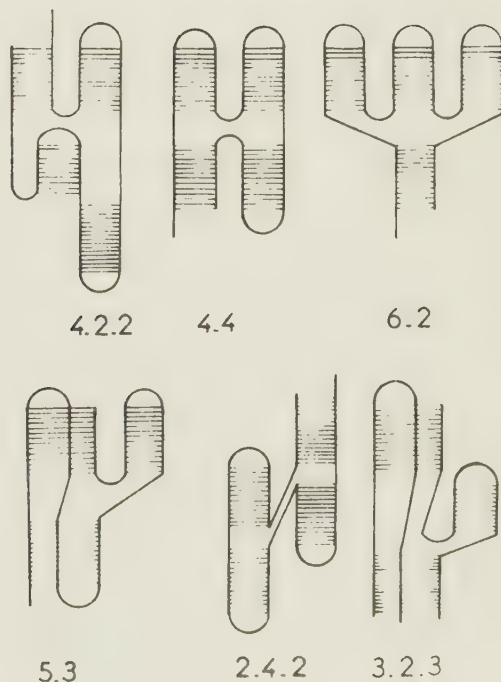


Fig I-a.
Séquence du tARN^{Val}_I de E. coli représentée dans l'arrangement en trèfle. La numérotation des résidus commence au pG terminal. La nomenclature des différentes régions de la chaîne est indiquée.

Fig. I-b.
Dénomination des modèles de tARN dans lesquels les régions hélicoïdales sont parallèles au grand axe de la molécule.



Malgré sa taille, l'adaptateur était compris au début, dans la perspective ouverte par Crick : une molécule dont le seul élément de spécificité était l'anticodon. Sa courbe de dénaturation thermique montrait qu'il pouvait être en double-hélice et son poids moléculaire indiquait qu'il contenait environ 75 nucléotides.

Les données chimiques montraient que l'acide aminé se fixait à une extrémité de la molécule, sur une séquence invariable de trois nucléotides.

Le modèle en épingle à cheveux suffisait largement à la compréhension du tARN. L'enzyme d'activation reconnaissait l'anticodon situé sur l'unique boucle de la molécule, et la partie terminale, flexible, permettait le rapprochement des acides aminés portés par des tARN adjacents lors de la traduction. Le groupe de Wilkins⁽⁴⁾ obtenait bientôt des cristaux (qui, par la suite, s'avérèrent constitués d'ARN ribosomal dégradé) et qui apportaient un argument de poids en faveur du modèle en épingle à cheveux. Des confirmations supplémentaires étaient apportées par le groupe de Cantoni⁽⁵⁾, qui s'était livré à l'étude des produits de digestion enzymatique de tARN total ou partiellement purifié.

Cependant, différents travaux contribuèrent à compliquer l'image. Les courbes de dénaturation thermique biphasiques obtenues par Fresco⁽⁶⁾ sur des accepteurs partiellement purifiés indiquaient qu'à l'intérieur d'un tARN existaient plusieurs régions hélicoïdales, distinguables par leur contenu en paires G.C, et que ce contenu variait d'une espèce de tARN à une autre. Par diffusion centrale des rayons X, Witz⁽⁷⁾ montrait que le tARN était plus compact que l'épingle à cheveux. A ce stade de la réflexion, les modèles présents à l'esprit, à double région hélicoïdale^(7,8) ou en triple hélice⁽⁹⁾, comportaient un ou deux sites fonctionnels de plus que dans l'épingle à cheveux. Cela correspondait aussi aux progrès des biochimistes qui s'interrogeaient sur les éléments de spécificité dans la structure du tARN. Que reconnaissent les enzymes d'activation ? Quel est le site de fixation sur les ribosomes ? Quel est le rôle des bases méthylées ? On commençait à rapprocher l'extrémité

acceptrice de l'anticodon. Passée la période euphorique où le secret de la vie était attribué à la structure en double-hélice de l'ADN, on commençait à trouver qu'autour du tARN le mystère s'épaississait.

Quand les premières séquences de tARN furent établies, celle légèrement inexacte de Holley⁽¹⁰⁾, puis celles de Zachau⁽¹¹⁾ et de Madison⁽¹²⁾, il apparut clairement que la structure devait être plus compliquée que prévu. Une étude de diffusion centrale des rayons X accompagnée de constructions de modèles effectuée par Lake et Beeman⁽¹³⁾ semblait restreindre les arrangements spatiaux de la molécule à deux seulement : la forme 4-4 et la forme 6-2. La convention adoptée ici pour désigner les modèles de tARN est indiquée sur la figure 1-b.

Avec la possibilité de travailler sur des accepteurs purifiés, les recherches s'orientèrent vers la détection d'effets plus fins, du second ordre, et à leur interprétation fonctionnelle.

Diverses expériences mettaient en évidence la sensibilité de la structure du tARN à la présence d'ions magnésium^(14,15), ou à l'acylation⁽¹⁶⁾, par chromatographie⁽¹⁶⁾, ultracentrifugation⁽¹⁷⁾ échange d'hydrogène⁽¹⁸⁾, pouvoir optique rotatoire⁽¹⁹⁾, résonance magnétique nucléaire⁽²⁰⁾, etc. Des mutants étaient obtenus⁽²¹⁾. On admettait que dans un modèle 4-4 ou 6-2, certaines parties de la molécule interagissaient⁽¹²⁾, et que la structure native, très sensible aux conditions de milieu, correspondait à un arrangement très précis, rigidement défini, des différentes régions de la molécule.

A ces petites modifications structurales, correspondaient des fonctions diverses dont certaines d'apparition récente : dans la régulation de la synthèse des ribosomes⁽²²⁾, dans la modulation de la vitesse de traduction⁽²³⁾, dans la reconnaissance par les transférases⁽²⁴⁾. On se doutait aussi que la traduction du message génétique ne se faisait pas selon une loi d'appariements

strictement complémentaires entre le codon et l'anticodon⁽²⁵⁾, et il était proposé que la structure de la boucle de l'anticodon dans le tARN contribuait aux propriétés générales de la correspondance entre les triplets et les acides aminés⁽²⁶⁾.

L'impression se dégageait peu à peu que la molécule de tARN pouvait être aussi compliquée qu'une protéine globulaire. Crick généralisait cette idée en suggérant qu'aux origines de la vie, les fonctions actuelles de certaines protéines étaient remplies par des acides nucléiques⁽²⁷⁾.

Tandis qu'un peu partout dans le monde, le tARN commençait à se laisser cristalliser⁽²⁸⁻³³⁾, les biochimistes rivalisaient d'ingéniosité pour agir en des points précis de la séquence, soit en modifiant chimiquement un résidu déterminé⁽³⁴⁻³⁶⁾, soit en introduisant une ou deux coupures spécifiques dans la séquence⁽³⁷⁻⁴¹⁾. De la masse des résultats, parfois contradictoires, et de données assez convergentes apportées par les techniques biophysiques, une image quand même plus précise a commencé à apparaître, et le débat semble se circonscrire autour des modèles 4-2-2⁽⁴²⁾ et 2-4-2⁽⁴³⁻⁴⁶⁾.

— — — — —

En l'absence de données cristallographiques suffisamment précises, la diffusion centrale des rayons X a largement contribué à façonner les idées sur la structure tridimensionnelle du tARN. Mieux encore, l'intérêt de ce problème a conduit à modifier la perspective des travaux fondés sur cette technique.

La thèse présentée ici a été entreprise sous la direction de M. V. Luzzati. Les études ont commencé sur des mélanges de tARN dont on déterminait le spectre de diffusion centrale, dans les conditions de collimation linéaire infinie. On cherchait alors à rendre compte des courbes obtenues par des modèles géométriques simples comme l'ellipsoïde, le cylindre ou le bâtonnet. Les paramètres structuraux globaux : masse, rayon de gyration, facteur d'asymétrie concordaient bien avec ceux obtenus par Witz⁽⁷⁾.

Mais ce genre de renseignements, sans autre exploitation, n'apportait que peu de lumière sur les problèmes biologiques qui donnent au tARN son intérêt. En gros, on arrivait à l'image d'un cylindre de 80 Å. de haut pour 25 Å. de diamètre. L'arrangement en feuille de trèfle⁽¹⁰⁻¹²⁾ commençait à émerger des séquences de tARN et paraissait peu compatible avec une répartition cylindrique de la matière ; les volumes géométriques simples devenaient tout-à-fait insuffisants pour interpréter les résultats. Par ailleurs, Guschlbauer avait proposé un modèle à trois filaments⁽⁹⁾ qui pouvait bien rendre compte, à première vue, des données structurales, mais que contredisait l'arrangement en trèfle. Comme il n'est pas dit que cet arrangement décrit des contraintes de séquence simultanément réalisées sur toutes les formes du tARN, on a étudié la possibilité de relâcher certaines contraintes et arranger la molécule dans un cylindre contenant des régions à trois filaments. C'est ce problème qui nous a conduit à rechercher les règles les plus générales qu'on pouvait énoncer concernant la structure tridimensionnelle des acides nucléiques quand la régularité est moindre que celle de la célèbre double-hélice d'ADN.

Les résultats de cette étude des propriétés de topologie et d'orientation des acides nucléiques sont exposées au chapitre 4. De manière inattendue, cette étude a suggéré la plausibilité d'une modification du trèfle contenant des régions à trois filaments, le modèle 5-3. Pour en vérifier la compatibilité avec les données de rayons X, une molécule de tARN dans cette configuration a été reproduite avec les composants atomiques⁽⁴⁷⁾ C.P.K.

Des techniques de calcul en ordinateur ont été mises au point dans cette période, pour la simulation des modèles moléculaires, et pour la détermination des intensités théoriques (Ch. 2 et 3).

A la même époque, Lake et Beeman⁽¹³⁾ arrivaient à des déterminations assez voisines des paramètres globaux du tARN total de levure. Leurs données expérimentales s'étendaient plus loin que les nôtres et permettaient par conséquent une meilleure

interprétation de la structure à courte échelle. Au terme de leur étude, ils proposaient le modèle 6-2, puis le modèle 4-4.

La possibilité d'obtenir des renseignements à des angles plus grands que ceux exploités jusque là par la collimation linéaire infinie était déjà apparue dans notre travail, où les courbes expérimentales étaient reproductibles jusqu'à $2\theta = 6^\circ$ à 7° . Cependant, le type de collimation employé limitait sévèrement le contraste dans cette région, et ne permettait pas d'interprétation claire. Aussi, avons-nous cherché à modifier les conditions de collimation, et passer en collimation linéaire finie. Ceci nous a également permis de diminuer considérablement le volume irradié par les rayons X et donc de travailler avec des quantités de matière moindres. Les modifications de l'appareil de rayons X, les équations de la collimation et la méthode employée pour effectuer les corrections de collimation sont expliquées au Ch. 2.

Les constructions de modèles, puis leur simulation et le calcul des intensités théoriques par des méthodes souples et rigoureuses devaient montrer assez rapidement que pas plus les modèles 6-2 que : les modèles 4-4 n'étaient en accord avec l'expérience, sans permettre toutefois d'exclure les modèles 5-3. Cependant, il restait loisible d'imaginer que les caractéristiques de diffusion d'un modèle 4-4 pouvaient, par un effet de moyenne, disparaître dans les mélanges de tARN.

C'est grâce à des dons généreux de M. Yaniv que le travail a pu prendre un cours nouveau. L'étude du tARN^{Val} non chargé de E.coli a confirmé et précisé les résultats obtenus avec le tARN total. De plus, une forme chargée et particulièrement stable, le phénoxy-acetyl tARN^{Val} a pu être examinée. Si les paramètres structuraux globaux sont extrêmement voisins dans les deux cas, par contre, une différence importante d'intensité s'est manifestée dans les grands angles ; la comparaison du tARN total chargé et non chargé confirmait cette différence. L'ensemble des résultats expérimentaux est exposé au Ch. 6. L'interprétation du niveau de l'intensité diffusée aux grands angles s'avérait particulièrement délicate, et nécessitait une révision critique des méthodes de

calcul employées ailleurs (Ch. 2).

Tandis que les expériences de rayons X se poursuivaient, Yaniv, Favre et Barrel démontraient⁽⁴⁸⁾ que deux résidus du tARN^{Val}, éloignés par leur position dans la séquence étaient en fait très voisins dans l'espace. Ce résultat restreignait considérablement les arrangements tridimensionnels envisageables pour le tARN ; il concordait avec nos résultats en ceci qu'il excluait les modèles 4-4 et 6-2. Une nouvelle construction de modèles a été entreprise, avec Yaniv et Favre, permettant de faire la synthèse des données physiques et biochimiques (Ch. 7).

Cependant, si un modèle est utile pour l'interprétation de l'ensemble des données physico-chimiques, il faut aussi qu'il aille au-delà et permette d'englober plus de faits que ceux sur lesquels il est fondé.

Le tARN possède, de manière plus ou moins directe, deux fonctions de reconnaissance : celle pour le codon, et celle pour l'acide aminé. Je me suis penché sur la première⁽⁴⁹⁾, en m'appuyant sur des arguments physico-chimiques et des calculs de mécanique quantique accomplis par P. Claverie. En ce qui concerne la reconnaissance éventuelle de l'acide aminé (ou de l'enzyme d'activation), le modèle 4-2-2 proposé au terme de cette étude n'est pas aussi parlant qu'on aurait pu le souhaiter. Je tenterai quand même en conclusion de faire une suggestion.

CHAPITRE 2

TECHNIQUES DE RAYONS X

La diffusion centrale des rayons X est une technique assez récente et dont l'utilisation dans l'étude des macromolécules biologiques tend à se généraliser. A l'inverse de la cristallographie classique, elle permet l'étude des molécules en solution et donne des informations reliées de manière directe à leur structure. Les informations sont moins détaillées que celles obtenues dans l'étude des cristaux puisqu'en solution, les molécules prennent toutes les orientations.

L'appareil utilisé ici pour l'étude de l'ARN de transfert en solution diffère peu de celui décrit par Luzzati, Witz et Baro⁽⁵⁰⁾. Le compteur de Geiger a été remplacé par un compteur proportionnel. Le type de collimation employé était au départ la collimation linéaire infinie ; dans ces conditions, l'intensité théorique diffusée par une particule de structure connue a une forme simple (Luzzati, 1957)⁽⁵¹⁾ et il est possible d'en déduire des paramètres géométriques liés à la forme et à la dimension de la particule (Luzzati, 1958)⁽⁵²⁾.

La stabilité et la fiabilité de l'appareil tel qu'il est décrit dans la réf. 50 ont été beaucoup améliorés depuis, grâce surtout aux efforts de M. P. Saludjian. De plus, nous avons été conduits par la suite à changer le mode de collimation pour les raisons exposées dans l'introduction, et à passer à la collimation linéaire finie. Il n'y a pas dans ce cas de formule analytique simple liant l'intensité théorique à la forme de la particule, et il a été nécessaire de ramener les courbes expérimentales aux conditions de la collimation ponctuelle. Les modifications d'appareillage et la méthode employée pour faire les corrections de collimation sont exposées aux § 3 et 5.

§ 1 . NOTATIONS.

Dans les conditions de la mesure, l'échantillon est supposé homogène, isotrope, limité par deux plans parallèles. Les notations utilisées ici diffèrent peu de celles employées précédemment⁽⁷⁾.

s élément de longueur dans l'espace réciproque, en $\text{\AA}^{-1}$. On a $s = 2 \sin \theta / \lambda$.
 η nombre d'électrons contenus dans une tranche cylindrique de l'échantillon, perpendiculaire aux faces, de surface 1 cm^2 .

γ énergie cohérente recueillie par élément de surface de l'espace réciproque, diffusée par un électron recevant un faisceau de rayons X d'énergie unitaire.

ρ_e densité électronique du solvant en $\text{el}/\text{\AA}^3$

R rayon de gyration électronique de la particule immergée dans le solvant.

$\bar{v}$ et ψ , volumes partiels spécifiques en cm^3/g et en $\text{\AA}^3/\text{el}$.

$j_n(s)$ fonction normée de l'intensité diffusée (voir réf. 51).

Les autres notations seront introduites au moment de leur utilisation.

Enfin, outre les abréviations classiques relatives aux acides nucléiques (A,T,G, C, etc. pour désigner les bases du tARN), nous utiliserons les suivantes :

n.e. : nombre d'électrons, en général.

d.e. : densité électronique, en général.

Rg : rayon de gyration électronique ; Rgt : Rg transversal.

§ 2 . RAPPEL DES EQUATIONS DE LA COLLIMATION LINEAIRE INFINIE.

Les équations de la collimation ponctuelle conduisent à une intensité théorique de laquelle il est possible d'extraire plusieurs paramètres.

a) - Milieu à deux densités constantes.

Connaissant les densités électroniques (constantes) du solvant et du soluté, le volume partiel spécifique du soluté et sa concentration électronique, on peut en déduire la surface spécifique s/v du soluté. A faible concentration, l'expression de s/v est pratiquement indépendante des paramètres cités.

b) - Solution de particules globulaires identiques.

Ayant déterminé la courbe de diffusion centrale extrapolée à concentration nulle, l'examen de sa partie gaussienne près de l'origine permet de calculer le rayon de gyration R_0 . Si l'intensité est de plus déterminée à l'échelle absolue, on peut calculer le nombre d'électrons de la molécule.

c) - Solution de particules globulaires identiques réalisant un milieu à deux densités constantes.

On peut alors, de surcroît, calculer le volume de la particule hydratée, après extrapolation à concentration nulle.

Si on appliquait les formules de la collimation ponctuelle aux courbes obtenues en collimation linéaire infinie, on obtiendrait un Rg et un nombre d'électrons "apparents", mais bien réels en ce sens qu'ils décrivent des propriétés géométriques de la particule considérée⁽⁵²⁾.

En pratique, cependant, on utilise des équations qui permettent de déterminer les valeurs vraies des paramètres. Il faut pour cela calculer quelques intégrales.

Une fonction qui intervient dans ces calculs est $j^+(s)$ qui peut se définir quand $j_n(s)$ admet un développement asymptotique en $1/s^3$:

$$\begin{aligned} s^3 j_n(s) &\simeq a + \delta s^3 \\ j^+(s) &= j_n(s) - \delta \end{aligned}$$

Une autre fonction intervenant dans les calculs est :

$$\phi(s) = j^+(s) - j^+(0) e^{-\frac{4\pi^2}{3} R_a^2 s^2}$$

où R_a est le R_g apparent défini par l'application de la loi de Guinier à $j_n(s)$:

$$j_n(s) = j_n(0) \left[1 - e^{-\frac{4\pi^2}{3} R_a^2 s^2} + \dots \right]$$

A l'origine, il est évidemment indifférent d'employer $j^+(s)$ ou $j_n(s)$.

Les fonctions nouvellement définies permettent le calcul de R_0 , de N le n.e. de la particule, de V son volume et de s sa surface quand la concentration électronique c_e du soluté est connue :

$$R_0^2 = \frac{R_a^2 + \left[\frac{g\sqrt{3\pi}}{16\pi^4} \int_0^\infty s^{-4} \phi(s) ds \right] \left[j_n(0) R_a \right]^{-1}}{1 - \left[\frac{\sqrt{3\pi}}{2\pi^2} \int_0^\infty s^{-2} \phi(s) ds \right] \left[j_n(0) R_a \right]^{-1}}$$

$$N c_e (1 - \rho_0 \psi)^2 = 2 \sqrt{\frac{\pi}{3}} j_n(0) R_a - \frac{1}{\pi} \int_0^\infty s^{-2} \phi(s) ds$$

$$\frac{s}{V} = \frac{16 \pi^2 a}{\int_0^\infty 2\pi s j^+(s) ds}$$

$$V = \frac{2 \sqrt{\frac{\pi}{3}} j_n(0) R_a - \frac{1}{\pi} \int_0^\infty s^{-2} \phi(s) ds}{\int_0^\infty 2\pi s j^+(s) ds}$$

§ 3 . DISPOSITIF EXPERIMENTAL.

Nous ne revenons pas sur le dispositif expérimental utilisé pour la collimation linéaire infinie. Le système de fentes est décrit sur la fig. 2-a. Pour réaliser une collimation linéaire finie, un nouveau système de fentes a été ajouté, qui opère dans une direction perpendiculaire au faisceau d'une part et au système initial d'autre part (fig. 2b).

En ϕ_1 , près de la source de rayons X, on a placé une fente fixe, taillée dans une plaque de cuivre. En ϕ_2 , à la sortie de F_I , la longueur du faisceau est limitée par deux vis cylindriques tronquées. Enfin, en F_4 est placée une fente taillée dans une plaque de cuivre, pouvant se déplacer dans son plan, parallèlement aux lèvres de F_4 . C'est par ce déplacement qu'on mesure la fonction de forme $g(t)$ du faisceau, définie plus loin. Si on appelle respectivement a_1 et a_2 les ouvertures de ϕ_1 et ϕ_2 , d_1 et d_2 les distances indiquées sur la fig. 1c, a_3 et a_4 la zone de lumière et celle de pénombre sur le plan du récepteur (voir fig. 1c), un calcul d'optimisation permet de relier certaines de ces quantités.

Soit r la résolution obtenue avec le système des fentes ϕ :

$$r \simeq a_3 + a_4/2$$

Posons $D = d_1 + d_2$

En prenant pour paramètres d_1 , D et r , et pour variable principale a_1 , il vient :

$$a_2 = \frac{d_1}{D} \left[r - a_1 \left(1 + \frac{D}{d_1} \right) \right]$$

$$a_3 = a_1 + (a_2 - a_1) \frac{D}{d_1}$$

L'intensité I reçue sur le plan du récepteur varie en gros comme le produit $a_1 a_3$. En écrivant que cette intensité est maximum, donc que la dérivée dI/da_1 est nulle, on trouve la condition :

$$a_1 = \frac{r d_1}{2D}$$

En pratique cependant, nous avons été limités par les possibilités du dispositif déjà en place.

Une autre modification des appareils existants, mais qui ne tient pas au changement du mode de collimation est que l'échantillon est maintenu fixe (auparavant, il tournait à une vitesse moitié de celle du goniomètre).

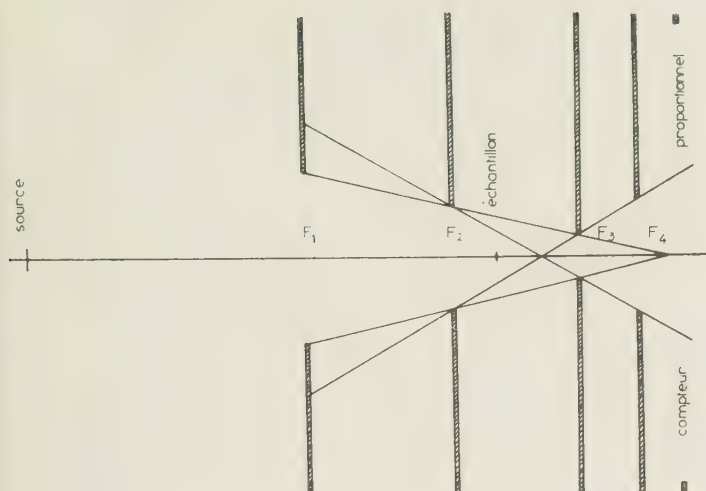


Fig. 2-a. Géométrie du système des fentes utilisé pour la collimation linéaire infinie. Le faisceau se dirige de la source vers le compteur proportionnel dans un plan perpendiculaire à celui de la figure. L'ensemble des fentes F_3, F_4 , et du compteur proportionnel, balaie l'espace réciproque autour d'un axe perpendiculaire au plan de la figure et qui passe par l'échantillon.

Fig. 2-b. Géométrie du système additionnel de fentes utilisé en collimation linéaire finie. Q_1, Q_2, Q_3 opèrent dans un plan perpendiculaire à celui relatif à F_1, F_2, F_3, F_4 .

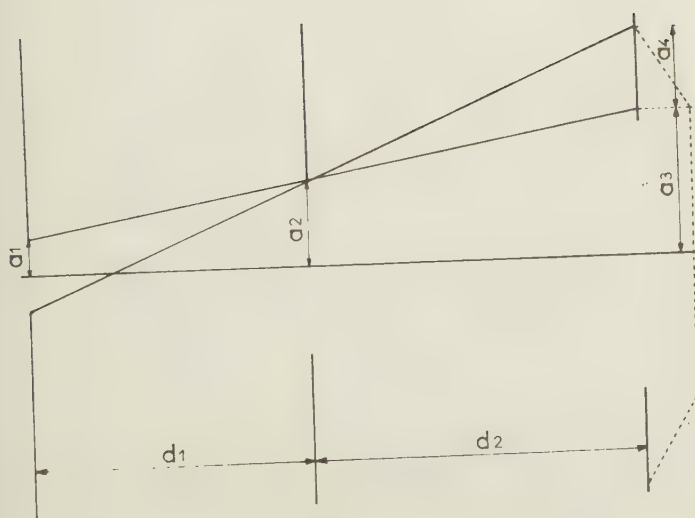
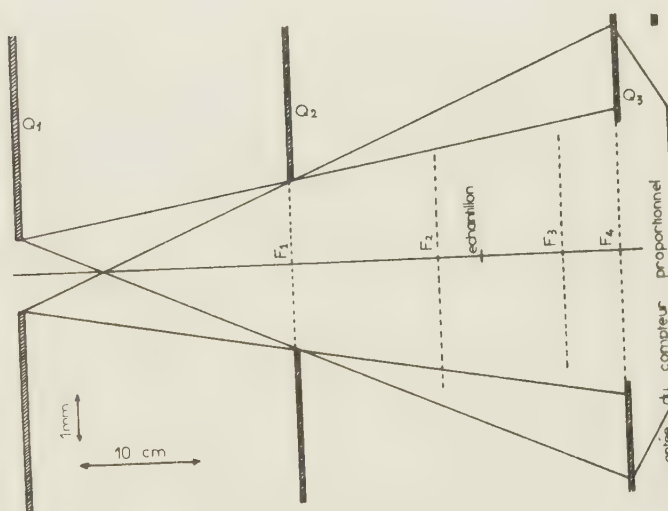


Fig. 2-c. Elements géométriques du calcul d'optimisation des ouvertures de fentes (voir texte).

Enfin, les résultats ont toujours été enregistrés sous forme digitalisée. Nous explicitons maintenant la manière dont les intensités mesurées avec un tel dispositif permettent de conduire aux intensités théoriques correspondant au cas de la collimation ponctuelle.

§ 4 . EQUATIONS DE LA COLLIMATION.

Considérons un faisceau ponctuel d'énergie g , qui coupe le plan R du récepteur en un point Q (Fig. 2d). Interposons au point E sur le trajet du faisceau un échantillon de matière diffusante. Dans le plan P, perpendiculaire en Q à la direction EQ du faisceau incident, l'intensité diffusée prendra la forme : $g I(\vec{QM})$, où on mesure $\vec{QM}$ en unités réciproques (A^{-1}). Soit M' l'intersection de EM avec le plan du récepteur. L'intensité reçue en M' par unité de surface réciproque est la même qu'en M : $g I(\vec{QM})$; pour l'instant nous assimilons une portion de sphère centrée en E à son plan tangent, P ou R suivant le cas.

Irradions maintenant l'échantillon par un faisceau de section arbitraire qu'on peut décomposer en faisceaux ponctuels élémentaires. L'intensité totale en un point M' du récepteur a en général une forme compliquée ; elle se simplifie beaucoup si on peut poser : $\vec{QM}' = \vec{QM}$. Ce cas se présente quand les faisceaux ponctuels élémentaires font avec Oz un angle toujours petit. Si $\delta\alpha$ est un tel angle, on a :

$$1 - \frac{QM'}{QM} \text{ de l'ordre de } (\delta\alpha)^2$$

Nous considérons dorénavant que cette approximation est réalisée ; il en résulte une conséquence très importante : deux faisceaux élémentaires d'énergie g_1 et g_2 , et coupant, en absence d'échantillon, R au même point Q, donneront au point M une intensité diffusée de la forme :

$$(g_1 + g_2) I(\vec{QM})$$

En d'autres termes, dans la limite où $\delta\alpha$ est petit, on ne se soucie pas de la direction du faisceau élémentaire, et ce qui importe est de connaître l'intensité totale du faisceau direct en chaque point du récepteur.

Soit $J(\vec{OM})$ l'intensité diffusée au point M par unité de surface, et $G(\vec{OQ}) dw$ l'intensité directe reçue autour du point Q sur un élément de surface (réciproque comme toujours) dw

$$J(\vec{OM}) = \int_{Q \in R} I(\vec{QM}) G(\vec{OQ}) dw$$

La mesure des intensités tant directe que diffusée, en pratique se fait sur un élément de surface assez important, qu'on ne peut traiter comme un infiniment petit. Par conséquent, tant $G(\vec{OQ})$ que $J(\vec{QM})$ sont connus de manière indirecte. Le produit de convolution de $G(\vec{OQ})$ ou $J(\vec{OQ})$ avec une fonction $S(\vec{OQ})$ décrivant la plage de sensibilité de l'appareil de mesure donne $G'(\vec{OQ})$ ou $J'(\vec{OQ})$, intensités directe et diffusée, mesurées dans les conditions de l'expérience.

La fonction qui nous intéresse est I . On pourrait dans un premier temps déduire G de G' et de S , J de J' et de S , puis I de J et de G . Ceci n'est pas nécessaire, car si

$$J(\vec{OM}) = \int_{Q \in R} I(\vec{QM}) G(\vec{OQ}) d\omega$$

on a aussi :

$$J'(\vec{OM}) = \int_{Q \in R} I(\vec{QM}) G'(\vec{OQ}) d\omega$$

En effet, le produit de convolution, dénoté ici X , est associatif.

Comme $J = I X G$ et que $J' = J X S$, on a $J' = (I X G) X S$ donc $J' = I X (G X S)$ et finalement $J' = I X G'$.

Soient deux axes rectangulaires, $\vec{Ot}$ et $\vec{Os}$ dans le plan R .

Tant en collimation linéaire finie qu'infinie, on s'est limité dans nos expériences aux faisceaux "très étroits" tels que $G'(s,t) = 0$ sauf sur l'axe $\vec{Ot}$, et on a cherché à mesurer $J'(s,0)$.

La formule :

$$J'(\vec{OM}) = \int_{Q \in R} I(\vec{QM}) G'(\vec{OQ}) d\omega$$

dans ces conditions devient :

$$J'(s,0) = \int_{-\infty}^{+\infty} I(-s,t) G'(0,t) dt$$

De plus, travaillant avec des solutions isotropes, I est centrosymétrique :

$$I(-s,t) = i(\sqrt{s^2+t^2})$$

En allégeant les notations, on arrive donc à une équation de la forme

$$j(s) = 2 \int_0^{\infty} i(\sqrt{s^2+t^2}) g(t) dt$$

et dont il faudra évidemment se garder d'oublier le sens et l'origine. Dans cette équation, la fonction théorique $i(s)$ est jusqu'à présent définie à une constante multiplicative près, de même que $g(s)$. On peut se donner de manière arbitraire une façon de normer $i(s)$ et $g(s)$, mais il faudra ensuite adopter pour la présentation des résultats expérimentaux sous la forme $j(s)$, une normation conséquente.

Normation.

Nous considérons maintenant que $j(s)$ est l'énergie diffusée dans les conditions de la collimation linéaire par un échantillon recevant le faisceau incident d'énergie $g(t)$. En collimation ponctuelle, l'énergie diffusée par le même échantillon a la forme: $i(s) \cdot g(0)$.

Dans ce traitement, l'intensité i est sans dimension, et j est une énergie. A partir de $i(s)$, on peut constituer la fonction normée $i_n(s)$:

$$i_n(s) = \frac{i(s)}{\eta^r}$$

qui est identique à la fonction $i_n(s)$ de la réf. 51.

Par ailleurs, on constitue à partir de $j(s)$ la fonction normée $j_n(s)$:

$$j_n(s) = \frac{j(s)}{\eta^r g(0)}$$

elle aussi identique à celle de la réf. 51.

De plus, on pose :

$$G(t) = g(t) / g(0)$$

Il en résulte que :

$$j_n(s) = 2 \int_0^{\infty} i_n(\sqrt{s^2 + t^2}) G(t) dt$$

En pratique, par conséquent, on mesure une énergie diffusée $j(s)$ et une énergie incidente à l'origine $g(0)$, puis on utilise les constantes physiques η et r pour passer à la fonction normée $j_n(s)$. Ayant obtenu la forme du faisceau incident, décrite par $G(t)$, on remonte à $i_n(s)$.

Pour un échantillon isotrope de volume V et de densité électronique moyenne ρ , $i_n(s)$ est liée de manière simple à la fonction théorique $I(s)$:

$$i_n(s) = I(s) / V\rho$$

$$s I(s) = 2 \int_0^\infty r p(r) \sin(2\pi r s) dr + P(0)$$

où $p(r)$ est la valeur moyenne de la fonction de Patterson $P(\vec{r})$ pour toutes les directions de $\vec{r}$, à module de $\vec{r}$ constant.

Inversement, partant d'un modèle de molécule bien défini, on peut calculer $p(r)$ puis $I(s)$ et de là, passer aux fonctions expérimentales $i_n(s)$ et $j_n(s)$.

§ 5 . CORRECTIONS DE COLLIMATION.

Connaissant g et j dans l'équation :

$$j(s) = 2 \int_0^\infty i(\sqrt{s^2 + t^2}) g(t) dt$$

il est possible d'en déduire i par des méthodes soit analytiques⁽⁵³⁾, soit numériques. Les méthodes analytiques utilisent en général les propriétés de la transformée de Fourier (voir par ex. Hossfeld⁽⁵⁴⁾). Cependant, nous nous sommes intéressés aux méthodes numériques en raison d'un algorithme très simple proposé par Lake⁽⁵⁵⁾. Son principe est le suivant : on se donne une fonction d'essai $E_1(s)$ qu'on substitue à $i(s)$ dans l'équation ci-dessus, et on appelle $T_1(s)$ le résultat :

$$T_1(s) = 2 \int_0^\infty E_1(\sqrt{s^2 + t^2}) g(t) dt$$

La comparaison de T_1 avec j renseigne sur la manière dont on peut modifier E_1 pour la rapprocher de i . On peut alors construire une seconde fonction d'essai E_2 , plus proche de i . Dans l'algorithme proposé par Lake,

$$E_{n+1}(s) = E_n(s) \cdot j(s) / T_n(s)$$

Au bout de n itérations, n étant voisin de la dizaine, on observe que $j - T_n / j$ est plus petit que 1% pour la plupart des valeurs de s . Il arrive cependant de trouver des valeurs de s telles que la convergence de $j - T_n / j$ vers zéro ne soit pas assurée, et c'est presque toujours le cas des petites valeurs de s , de telle sorte que la détermination du rayon de gyration par cet algorithme est quasiment impossible.

Nous avons choisi un algorithme légèrement différent, dans lequel le rapport $E_{n+1}(s) / E_n(s)$ tient compte de toutes les valeurs de u pour lesquelles $E_n(u)$ a

contribué à $T_n(s)$.

Plus précisément, si $j(s)/T_n(s) - I \neq 0$, il existe au moins une valeur de u telle que :

$$\begin{cases} \mu \geq s \\ (i(u)/E_n(u) - 1)(j(s)/T_n(s) - 1) > 0 \\ g(\sqrt{u^2 - s^2}) \neq 0 \end{cases}$$

et la contribution de cette valeur de u à l'écart entre $j(s)$ et $T_n(s)$ est d'autant plus importante que $g(t)$ est grand ($\sqrt{s^2 + t^2} = u$). D'où la formule théorique proposée pour le passage de E_n à E_{n+1} :

$$E_{n+1}(\mu) = E_n(u) \frac{\int_0^\mu \frac{j(s)}{T_n(s)} g(t) dt}{\int_0^\mu g(t) dt}$$

$$(s = \sqrt{u^2 - t^2})$$

qui s'interprète ainsi : l'écart entre $i(u)$ et $E_n(u)$ se reflète avec une "probabilité" $g(t)$ dans l'écart entre $j(s)$ et $T_n(s)$.

La méthode de Lake revient à substituer dans cette formule la fonction g par une fonction δ à l'origine. Dans le cas de nos calculs, nous avons remplacé dans la formule ci-dessus g par une fonction formée de deux segments linéaires et qui approximativement sa forme.

De cette manière, la convergence est satisfaisante et la limitation dans ^{la} précision provient uniquement de la grandeur du pas choisi pour l'intégration. A l'origine notamment, on ne saurait se satisfaire d'un pas de 10^{-3} A^{-1} .

Nous avons choisi de remplacer les premiers points de $E_n(s)$ par une exponentielle de la forme e^{-ks^2} , ceci jusqu'à $s = 8 \text{ à } 10 \cdot 10^{-3} \text{ A}^{-1}$, et de choisir un pas d'intégration de 10^{-4} A^{-1} . Evidemment, l'espacement entre les points expérimentaux est plus grand : $5 \cdot 10^{-4} \text{ A}^{-1}$ à l'origine, 10^{-3} A^{-1} un peu plus loin et jusqu'à $s = 20 \cdot 10^{-3} \text{ A}^{-1}$, et $3 \cdot 10^{-3} \text{ A}^{-1}$ au-delà. En dehors de la région exponentielle, $E_n(\sqrt{s^2 + t^2})$ est déterminé par une interpolation linéaire entre $E_n(s_1)$ et $E_n(s_2)$ où s_1 et s_2 représentent les deux nombres entiers de fois 10^{-3} A^{-1} qui encadrent $\sqrt{s^2 + t^2}$. Le fait que la loi de Guinier soit vraie à l'origine justifie l'emploi d'une "interpolation exponentielle" précise à l'origine.

§ 6 . CALCUL DES INTENSITES THEORIQUES.

Il est facile de calculer l'intensité diffusée par une particule dans le vide, mais si la particule est en solution, le problème se complique, même quand on suppose, ce que nous faisons par la suite, que la densité électronique (d.e.) du solvant est uniforme. Dans ces conditions, les interactions entre solvant et soluté se réduisent à l'exclusion du solvant hors du volume occupé par le soluté ; les effets de solvation qui peuvent se produire à la surface du soluté sont négligés.

Pour commencer, représentons chaque atome par une fonction delta multipliée par le nombre d'électrons (n.e.) porté par l'atome. La distribution de la d.e. d'une particule en solution est alors décrite par :

$$\rho(\vec{r}) = \sum_i n_i \delta(\vec{r} - \vec{r}_i) - \rho_0 V(\vec{r}) \quad (1)$$

où n_i et $\vec{r}_i$ définissent le n.e. et la position du ième atome de la particule, où ρ_0 est la d.e. (uniforme) du solvant, et $V(\vec{r})$ une fonction qui vaut 1 à l'intérieur du volume occupé par la particule et zéro au-dehors : appelons-la la fonction de volume.

De là, on passe à la fonction de Patterson :

$$P(\vec{u}) = \iiint_{\vec{r}} \rho(\vec{r}) \rho(\vec{r} + \vec{u}) d\vec{r}$$

puis à $i(s)$ d'après la formule du § 4.

Si de plus on nomme V le volume de la particule, ψ le volume partiel spécifique en $\text{\AA}^3$ par électron et $N = \sum n_i$ le n.e. total, on a :

$$i(0) = \left(\sum_i n_i - \rho_0 V \right)^2 = N^2 (1 - \rho_0 \psi)^2$$

Quand s croît, et qu'il n'y a plus de corrélation de phase entre les amplitudes diffusées par chaque atome, $i(s)$ tend vers une valeur asymptotique :

$$\text{val. asympt. de } i(s) = \sum_i n_i^2$$

Le calcul numérique de $P(\vec{r})$ est facilité quand on peut décrire $V(\vec{r})$ par une fonction analytique simple (ellipsoïde, sphère...) ; mais pour nos besoins où le TAR^N est défini par un ensemble d'atomes, il faut trouver un critère qui permette de dire si un point est à l'intérieur ou à l'extérieur de la molécule. Une méthode, dite "méthode des atomes effectifs", proposée par Langridge⁽⁵⁶⁾ tourne cette difficulté. C'est elle que nous avons adoptée dans les premiers calculs.

Dans la formule (I) on peut découper $V(\vec{r})$ en n fonctions de volume $V_i(\vec{r})$,

chacune associée à un atome et décrivant son volume de Van der Waals dans l'état chimique où il se trouve. On peut alors, dans $\rho(r)$ regrouper les deux termes du deuxième membre :

$$\rho(\vec{r}) = \sum_i \left[n_i \delta(\vec{r} - \vec{r}_i) - \rho_0 V_i(\vec{r}) \right]$$

Pour se débarrasser des $V_i(\vec{r})$ et traiter la molécule comme un ensemble de points, Langridge est conduit à utiliser ρ comme une fonction de deux variables :

$$\rho(\vec{r}, s) = \sum_i n'_i(s) \delta(\vec{r} - \vec{r}_i) \quad (2)$$

dans cette expression, $n'_i(s)$ est une fonction empirique appelée nombre d'électrons effectifs de l'atome. Elle est construite de telle manière que l'intensité théorique diffusée par un atome isolé soit identique, qu'on utilise la formule (1) ou la formule (2).

$n'_i(0)$ a une définition simple : il est égal à la différence entre le n.e. de l'atome et le n.e. du solvant exclu du volume v_i occupé par cet atome dans la molécule. D'autre part, quand s tend vers l'infini, $n'_i(s)$ tend vers n_i , le n.e. de l'atome dans le vide.

Avec cette méthode, $i(0)$ a la valeur correcte :

$$i(0) = \left(\sum_i n'_i \right)^2 = N^2 (1 - \rho_0 \psi)^2$$

En principe, également, $i(s)$ devrait tendre vers la valeur asymptotique correcte, puisque pour un atome isolé, $n'_i(\infty) = n_i$.

On peut cependant faire deux critiques à ce procédé.

La première est que le volume v_i , qui devrait être une propriété de l'atome dépend en fait de la configuration de la molécule (sinon, il serait possible de calculer tout volume partiel spécifique comme une somme de contributions atomiques, ce qui, à 10% près, n'est pas encore le cas). Elle rend mal compte des trous qui peuvent exister au voisinage d'une paire de liaisons hydrogène dans une structure en double-hélice. Ou alors, il faudrait ajuster les coefficients empiriques à chaque cas particulier, ce qui retire à la méthode, une partie de sa valeur.

La seconde, plus profonde, est que la méthode revient à soustraire non pas un solvant de d.e. uniforme, mais un solvant dont les molécules ont mêmes centres que les atomes de la particule, ce qui a pour effet d'estomper les fluctuations à courte distance de la d.e.

Nous avons utilisé cette méthode, pour l'interprétation des résultats en collimation linéaire infinie. Dans ce cas, une étape du calcul est la soustraction de

l'intensité diffusée à $\pm 8^\circ$; les courbes expérimentales ne vont pas au-delà et l'imperfection de la méthode est peu sensible.

Les premiers résultats des expériences effectuées en collimation linéaire finie nous ont conduit à remettre en question cette façon de calculer les intensités, et nous avons mis au point une méthode originale, dite méthode des cubes que nous allons maintenant décrire.

Considérons un volume de forme quelconque, mais assez grand pour contenir complètement la particule, et divisons-le en cellules cubiques. Nous décrirons la particule, non plus par ses atomes, mais par les positions des centres des cubes affectés de facteurs de structure convenablement choisis : le coefficient est nul si le cube ne contient pas d'atome (de centre d'atome, pour être plus précis), sinon il est pris égal à la différence entre le n.e. total des atomes contenus dans le cube, et le n.e. du solvant occupant un même volume.

Cette manière de décrire la molécule substitue à la d.e. réelle ρ une d.e. ρ_2 qui s'en déduit par des opérations mathématiques simples et bien définies :

$$\rho_1(\vec{r}) = \rho(\vec{r}) \times v(\vec{r})$$

$$\rho_2(\vec{r}) = \sum_{i,j,k} \rho_1(\vec{r}) \delta(\vec{r} - i\vec{u} - j\vec{v} - k\vec{w})$$

$\rho \times v$ est le produit de convolution de la fonction ρ par la fonction de volume relative au cube ; $\vec{u}$, $\vec{v}$ et $\vec{w}$ sont des vecteurs orientés le long des axes du réseau des cubes et de longueur égale aux côtés des cubes. Pour alléger cette expression, numérotions les cubes de $i = 1$ à m et appelons $\vec{a}_i$ le vecteur qui définit le centre du i ème cube. Il vient :

$$\rho_2(\vec{r}) = \sum_{i=1}^m \rho_1(\vec{r}) \delta(\vec{r} - \vec{a}_i)$$

ρ_2 se déduit de ρ_1 par un échantillonnage aux noeuds du réseau cubique constitué par les centres des cubes.

Si on appelle F_1 , F_2 et F les transformées de Fourier de ρ_1 , ρ_2 et ρ , on a :

$$F_2(\vec{s}) = \sum_i F(\vec{s} + \vec{a}_i^*)$$

où les vecteurs $\vec{a}_i^*$ désignent les positions des noeuds du réseau réciproque[†]. Pour $|\vec{s}|$ petit devant $|\vec{a}_j^*|$, la contribution du terme $F(\vec{s} + \vec{a}_j^*)$ est petite puisque $F(\vec{s})$ décroît rapidement. Nous ne tiendrons compte que d'une seule contribution, celle où $\vec{a}_j^* = 0$ d'où :

$$F_2(\vec{s}) = F_1(\vec{s})$$

[†] a^* désigne le complexe conjugué de a

On en déduit alors F_2 :

$$F_2(\vec{s}) = F(\vec{s}) \cdot \varnothing(\vec{s})$$

où $\varnothing(\vec{s})$ est la transformée de Fourier de $v(\vec{r})$.

L'expression de l'intensité devient :

$$i_2(s) = \langle I(\vec{s}) \cdot \varnothing^2(\vec{s}) \rangle \propto i(s) \cdot \varphi^2(s),$$

en faisant l'hypothèse qu'à $|\vec{s}|$ constant, les variations de $I(\vec{s})$ et $\varnothing(\vec{s})$ sont indépendantes.

Nous pouvons maintenant tenir compte de la forme des atomes, au moins de manière approchée, par un coefficient qui décrit la variation moyenne des facteurs de structure atomique. Nous choisissons le facteur de structure f_N de l'azote comme représentatif de l'ensemble. Il vient finalement :

$$i(s) = \frac{i_2(\vec{s})}{\varphi^2(\vec{s})} \cdot \frac{f_N^2(s)}{f_N^2(0)}$$

La fonction $\varphi^2(s)$ a été tabulée⁽⁵⁷⁾ ; $f_N(s)$ peut être trouvée dans les tables internationales de cristallographie : l'emploi de cette méthode ne soulève pas de difficulté majeure.

Quoique d'un point de vue mathématique, le volume v soit arbitraire, on est obligé de le choisir voisin du volume occupé par une molécule de solvant.

Si la taille choisie pour le cube est trop grande, la description de la molécule par des cubes risque de s'étendre loin au-delà de sa surface et cela revient à l'entourer d'une couche de densité électronique bien inférieure à celle du solvant. Si au contraire, la taille du cube est trop petite, la molécule sera décrite par des portions de volume localisés autour de ses atomes, et cela équivaut à considérer par exemple que le solvant s'insère en des lieux où normalement il ne peut avoir accès.

Dans la pratique, nous avons choisi 3 Å pour côté du cube ; un tel cube contient en moyenne 9 des dix électrons d'une molécule d'eau. On a également étudié la variation du Rg de l'image de la particule en fonction du côté du cube. Elle est lente jusqu'à 3 Å. Ensuite, le Rg chute rapidement.

La méthode décrite se prête bien au calcul en ordinateur. Elle est économique puisque le nombre d'atomes composant la molécule est réduit, et que dans le calcul de la fonction de Patterson, il n'est pas nécessaire d'introduire une variation du facteur de structure effectif avec l'angle : la correction se fait au stade des intensités. La description de la molécule par son image dans un réseau cubique est aisée, elle utilise une correspondance entre un système de trois coordonnées et un indice variant dans un

tableau à une dimension, s'appuyant sur la théorie des anneaux d'intégrité.

Il aurait été intéressant de disposer de données expérimentales sûres et déjà interprétées de manière définitive, pour éprouver la validité des méthodes de calcul. Cela n'est pas encore possible. Les courbes de diffusion centrale obtenues sur l'ADN par Kirste et Oberthür⁽⁵⁸⁾ d'une part, par Bram et Beeman⁽⁵⁹⁾ d'autre part diffèrent justement dans la région des grands angles où l'écart est le plus grand entre les intensités calculées par la méthode des cubes et par celle de Langridge. La courbe théorique calculée par la méthode des cubes donne une intensité diffusée aux grands angles inférieure à celle déterminée par Bram et Beeman, et légèrement supérieure à celle déterminée par Oberthür et Kirste (Fig. 2-e). La courbe théorique calculée par la méthode de Langridge prévoit une intensité théorique qui est inférieure aux deux intensités expérimentales. Nous reviendrons sur ces résultats lors de l'interprétation de l'intensité diffusée aux grands angles par le tARN.

§ 7 . DEPOUILLEMENT DES DONNEES.

Pour augmenter la précision des mesures, un soin tout particulier a été porté au relevé des données et à la détermination des différentes constantes accessoires intervenant dans la normation de l'intensité diffusée. Au début, les courbes étaient établies après lissage d'un enregistrement graphique. Par la suite, on a adjoint un enregistrement digitalisé qui a permis à précision égale, un gain dans le temps de dépouillement appréciable. Pour chaque cuve, le facteur d'absorption des lames de mica était mesuré ainsi que, lorsque cela était nécessaire, leur diffusion aux très petits angles, ou aux grands angles. Dans les expériences en collimation linéaire finie, la diffusion des micas était systématiquement établie sur toute la région des mesures.

L'épaisseur des cuves était mesurée sur un banc d'optique à l'aide d'un système de deux microscopes. La transmission d'une cuve est une fonction de l'épaisseur l donnée par la loi de Beer :

$$T = e^{-k_a l}$$

où k_a dépend de la nature de l'échantillon. Pour le tARN dans l'eau, aux concentrations employées (0,3 à 5%), à 0,1M NaCl on peut écrire :

$$(\log_{10} T)/l = 0,431 + 0,054 \cdot c$$

c étant la concentration du tARN en g/100ml.

Pour chaque série d'expériences, on déterminait la loi de k_a et on calculait l'épaisseur d'après le facteur de transmission.

Les opérations qui conduisent des résultats bruts aux intensités normées sont simples mais fastidieuses. Il faut, tenant compte des filtres utilisés et de la vitesse

de balayage, normer les faisceaux directs ; il faut normer la diffusion par l'air et la soustraire de toutes les expériences ; il faut calculer les épaisseurs de chaque cuve, normer le solvant et le soustraire de toutes les expériences relatives au soluté, en tenant compte du volume partiel spécifique : si on a une solution à 3% d'un matériel de volume partiel spécifique de $0,5 \text{ cm}^3/\text{g}$ la contribution réelle du solvant est égale à 0,98 fois celle du solvant normé. Il faut enfin normer les courbes à concentration électronique unitaire.

Tous ces calculs sont maintenant très aisément accomplis sur ordinateur grâce à un programme dont la structure permet une grande souplesse d'utilisation.

CHAPITRE 3

MODELES MOLECULAIRES.

§ I . GENERALITES SUR L'EMPLOI DES MODELES MOLECULAIRES.

Ces dernières années, l'emploi des modèles atomiques s'est généralisé en biologie moléculaire, devenant une technique de laboratoire courante. Cependant, depuis le succès spectaculaire de Watson et Crick dans la construction du modèle en double-hélice d'ADN, peu d'exemples probants peuvent être signalés.

Le travail de Browne⁽⁶⁰⁾ sur l' α -lactalbumine de boeuf fournit un exemple idéal d'emploi des modèles moléculaires. En effet, la séquence de l'enzyme considérée présente de grandes analogies avec celle du lysozyme dont la structure cristallographique est connue. La question s'est posée de savoir si la chaîne polypeptidique de l' α -lactalbumine adoptait une conformation voisine de celle du lysozyme. Une étude de modèles moléculaires a montré alors que la réponse pourrait bien être positive.

Souvent, les modèles moléculaires cherchent à pallier à l'insuffisance des données chimiques, plutôt que cristallographiques. C'est le cas des modèles d'interactions entre acides aminés et codons⁽⁶¹⁾, ou entre polypeptides et acides nucléiques⁽⁶²⁾.

Des études nombreuses ont été consacrées à la construction de modèles tridimensionnels d'ARN de transfert, utilisant les matériaux les plus divers : des plaques de métal et des ressorts⁽⁶³⁾, des blocs "Lego" d'un jeu de constructions pour enfants⁽⁶⁴⁾, des modèles atomiques vrais, squelettiques ou compacts.

Dans les modèles squelettiques, la molécule est représentée par un ensemble de tiges, dont chacune figure une liaison chimique entre deux atomes. Plusieurs tiges peuvent être préalablement groupées pour représenter les diverses familles d'atomes, ou les groupements fonctionnels caractéristiques (cycle aromatique, groupe amide, etc.) .

La famille des modèles compacts doit sa mise au point aux progrès conjoints de la chimie quantique et de la technologie des matières plastiques. La possibilité de représenter une molécule par un assemblage de boules, de tétraèdres, de quartiers de sphères, est une idée scientifique peu banale, dont le mérite revient en grande partie à Linus Pauling.

Ces dernières années, le marché des modèles compacts a été dominé par les composants C.P.K. (C pour Corey, P pour Pauling et K pour les connecteurs Koltun). Ces modèles sont légers, maniables, et reproduisent en principe les angles et distances de liaison avec leurs tolérances usuelles : la flexibilité mécanique simule les tolérances chimiques⁽⁴⁷⁾.

Il existe également une famille de modèles qui combine les propriétés des deux familles précédentes : des revêtements en mousse de plastique donnent la sensation de volume pour une molécule construite avec des composants initiaux squelettiques.⁽⁶⁵⁾

Dans le présent travail, la construction de modèles s'est conjuguée aux calculs effectués en ordinateur. Disposant de données expérimentales, on cherchait une structure qui permettrait d'en rendre compte. L'ordinateur pourrait, en principe, moyennant certaines hypothèses, essayer toutes les structures possibles une à une, mais toute la difficulté est justement dans le choix des hypothèses, et l'appréciation des divers facteurs. L'absence de données expérimentales variées dans le domaine des structures tertiaires d'acides nucléiques n'autorisait la formulation de règles empiriques que très générales et trop évidentes (complémentarité des appariements). Une tentative d'explicitation de certains principes de construction des acides nucléiques sera faite au chapitre 4.

§ 2 . CONSTRUCTION DE MODELES DE tARN.

Les constructions de modèles ont été entreprises très tôt, dès que la courbe de diffusion centrale du tARN total de levure eut été établie. On ne connaissait alors qu'une séquence de tARN et on ne disposait pas de données cristallographiques relatives aux ARN viraux. Il fallait donc au départ se faire une idée des divers arrangements secondaires envisageables, y compris ceux qui incluaient des portions à trois ou à quatre filaments.

Les premiers oligonucléotides ont été reproduits avec^{des} composants tout à fait rudimentaires, l'imagination et la théorie tentant de suppléer à leur imperfection.

Avec la disponibilité sur le marché des composants C.P.K., puis la parution des coordonnées atomiques relatives aux ARN viraux en double-hélice, le travail a pris un cours nouveau.

Les études menées par le groupe du King's College de Londres⁽⁶⁶⁻⁷⁰⁾ montrent que les ARN viraux en double-hélice ont une structure voisine de celle de l'ADN dans la forme⁽⁷¹⁾ A. Le Rgt est grand : 8,5 Å environ ; l'axe de l'hélice passe

en-dehors des bases ; le pas de l'hélice est voisin de 30 Å , et le nombre de bases par tour est de 10 ou 11^(66,67) ou 12⁽⁶⁸⁻⁷⁰⁾ .

Nous nous sommes inspirés des coordonnées de l'ARN 10 ou 11 pour construire les segments ordonnés. Avec les modèles C.P.K., les bases puriques et pyrimidiques se construisent sans difficulté, de même que la chaîne des phosphates. Par contre, la possibilité de reproduire exactement la configuration du ribose est exclue. Le cycle du pentose, qui comprend un oxygène divalent et quatre carbones tétraédriques n'est pas plan. Tout comme le cycle des hexoses avec ses configurations en chaise ou en bateau, celui des pentoses adopte des configurations variées qui s'énoncent ainsi : Cn' endo ou Cn' exo. Dans cette notation, le prime indique que le carbone appartient au sucre et non à la base ; n représente le numéro de l'atome de carbone du cycle qui n'est pas dans le plan formé par l'oxygène et les trois autres carbones ; le qualificatif endo ou exo indique si le carbone hors-cycle est orienté vers la liaison glycosidique, ou bien vers l'extérieur.

Tous les riboses ont été uniformément construits dans une configuration C2' endo , qui est celle du désoxyribose dans l'ADN . Une compilation récente^(72,73) montre que dans les ribonucléotides, de même que dans les ARN viraux, la configuration C3' endo prévaut. Au degré de résolution des expériences de diffusion centrale des rayons X, la configuration précise du sucre n'est évidemment pas vue. Lors de la simulation en ordinateur des modèles en vue du calcul de la fonction de Patterson, les segments ordonnés construits étaient de toutes façons remplacés par des sections en double-hélice conformes aux données cristallographiques.

Si, dans la construction des régions ordonnées, il est commode de prendre pour guide les coordonnées cristallographiques des ARN viraux, on dispose de peu de renseignements pour construire les boucles. Les études sur la réactivité de divers résidus du tARN avec des agents chimiques, et celles sur la susceptibilité aux attaques enzymatiques mènent à la conclusion générale que les boucles ont une disposition spatiale bien définie, et par conséquent excluent des structures complètement étirées.

Une boucle qui a fait l'objet d'une attention particulière est celle de l'anticodon. Fuller et Hodgson⁽²⁴⁾ ont avancé une argumentation plausible mais non convaincante pour proposer que dans cette boucle, cinq bases sont empilées dans le prolongement d'un des filaments du tronçon hélicoïdal auquel la boucle est raccordée, tandis que les deux autres bases (des pyrimidines) sont orientées vers l'extérieur. Ce modèle essayait de donner une interprétation mécaniste de l'hypothèse du wobble de Crick et de ce fait, dans de nombreux articles, il a été admis que la structure de la boucle de l'anticodon avait été "établie" par Fuller et Hodgson.

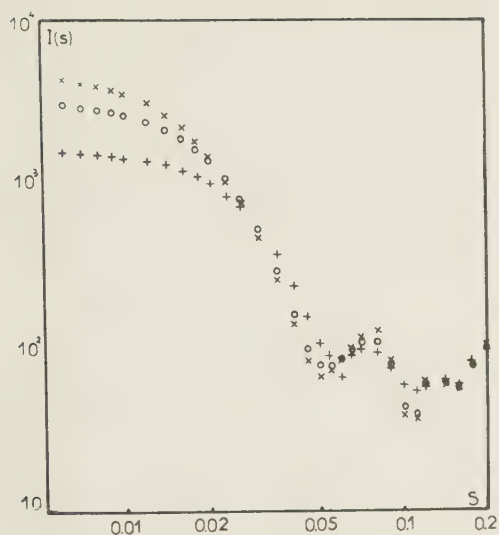


Fig. 3-a .

Influence de la longueur de chaîne : un segment en double-hélice d'ARN de forme IO de longueur variable. + : 5 paires de bases ; o : 10 paires ; x : 15 paires . On voit que l'écart entre les deux plateaux aux grands angles croît avec la longueur de l'hélice. Il se stabilise vers 15 plateaux.

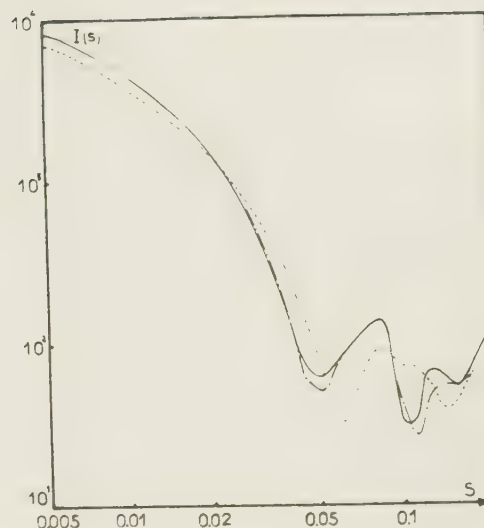


Fig. 3-b.

Influence de la géométrie locale : un segment en double-hélice de 30 paires de bases, avec sodium dans diverses configurations. --- ADN forme B ; — ARN forme IO ; -.- ARN forme II .

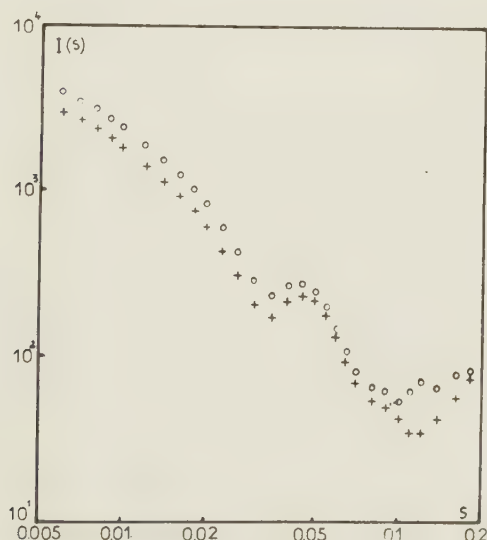


Fig. 3-c.

Influence de la nature chimique de la molécule. Un segment à un filament, de 30 nucléotides de cytosine (+) ou de guanine (o) dans la configuration de l'ARN IO . Dans les hélices à deux filaments, la substitution d'une paire A.U à une paire G.C a un effet négligeable. L'addition de sodium à l'extérieur de la molécule donne un effet visible mais restreint.

En fait, comme certaines prémisses de l'hypothèse du wobble peuvent être critiquées (Chapitre 8), nous ne nous sommes pas limités, dans la construction des boucles, à ce modèle dont la validité, par ailleurs, est assez indépendante de la théorie à laquelle il veut donner un fondement.

Plusieurs sortes de boucles ont été construites. Parmi les boucles à sept nucléotides une était conforme au modèle de Fuller ; une autre, un peu moins compacte, avait ses trois bases médianes dirigées vers l'extérieur ; une troisième, enfin, était très allongée.

Après simulation en ordinateur, il apparaissait en général que le rayon de gyration était plus petit que le R_g expérimental, et par conséquent on était conduit à substituer le second type de boucles au premier ; le troisième, jugé peu physique a été très rarement utilisé.

Dans les boucles à plus de 7 nucléotides représentant la boucle DiHU, on s'est servi le plus souvent d'une boucle semblable à la deuxième, ou, pour les besoins des modèles 4-2-2, d'unités construites "à la demande". En effet, dans ces modèles, la région de l'espace que peut occuper la boucle est sérieusement limitée par l'encombrement stérique des régions voisines.

Outre la construction des modèles 4-2-2 (Chapitre 7), des modèles de différentes natures ont été étudiés, et les intensités théoriques qui leur correspondent ont été établies.

a) Modèles 6-2.

Lake et Beeman⁽¹³⁾ ont été les premiers à en proposer. Leur argument était fondé sur leurs expériences de diffusion centrale des rayons X. L'intensité théorique qu'ils obtenaient pour des modèles en trèfle ouvert présentait une épaule vers $s = 0,03 \text{ \AA}^{-1}$ qu'ils interprétaient comme reflétant la participation, dans la fonction de Patterson, des distances entre atomes appartenant à des bras différents. Or, l'intensité diffusée par un solide formé de deux sphères présente une telle épaule, qui disparaît quand les sphères viennent en contact. Lake et Beeman estimèrent que l'absence d'épaule dans la courbe expérimentale s'expliquait si le maximum de contacts existait entre les bras du tARN, d'où leur proposition, qui n'est d'ailleurs pas un "modèle" en ce sens qu'ils n'ont pas présenté de dessin indiquant le contour possible de la chaîne polynucléotidique.

Leur idée a été reprise par la suite⁽⁶⁴⁾ par Cramer et son groupe, qui se sont appuyés sur des données chimiques suggérant que tel ou tel résidu, étant inaccessible à certains réactifs chimiques, devait être engagé dans une paire A.U, G.C ou G.U. Ils se sont également appuyés sur des considérations relatives aux corrélations de séquence dans les tARNs.

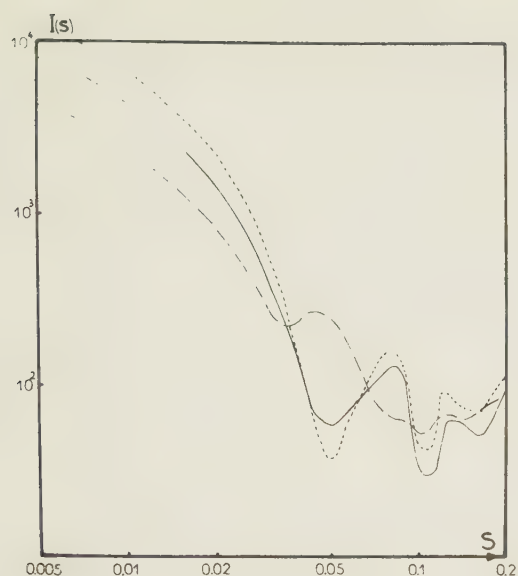


Fig. 3-d. Influence du nombre de brins. Un segment de 30 plateaux de nucléotides dans la géométrie de l'ARN IO, à 1, 2 ou 3 filaments.

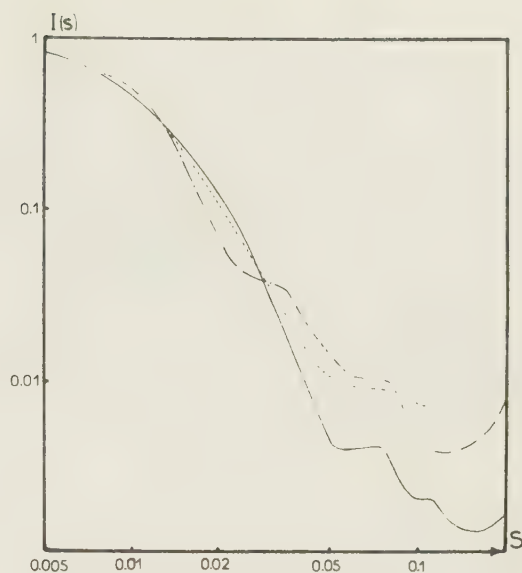


Fig. 3-e. Un modèle 6-2 (— · — ·), comparé au tARN^{Val} chargé (---) et non chargé (—).

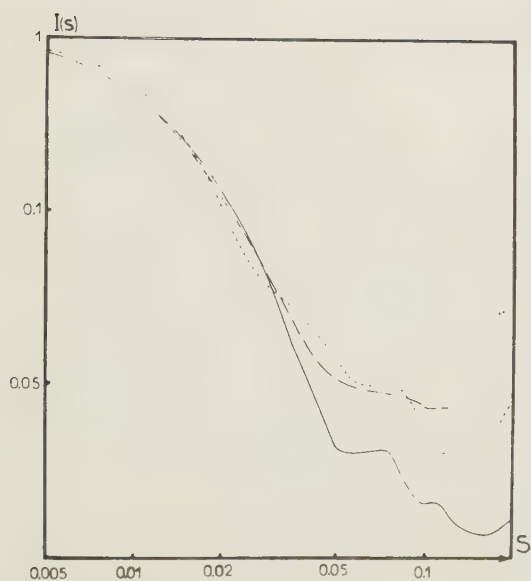


Fig. 3-f. Un modèle 4-4 faisant 23 Å de R_g comparé au tARN^{Val} chargé et non chargé.

Divers modèles 6-2 ont été construits avec les composants C.P.K. ; il est apparu rapidement que ces modèles étaient en contradiction avec les courbes expérimentales, tant les nôtres que celles de Lake et Beeman. En effet, si trois des bras de la molécule sont en contact étroit, le rayon de gyration est faible, de l'ordre de 18 Å. Mais si on sépare les bras pour augmenter le rayon de gyration, on a une structure lâche, sans interaction supplémentaire entre les boucles, donc ne répondant plus aux préoccupations de ceux qui l'ont proposée. De plus, les intensités théoriques (Fig. 3-e) présentent une épaule vers $s=0.03$ incompatible avec l'expérience.

b) Modèles 4-4.

Le premier de ces modèles a également été proposé par Lake et Beeman qui, après avoir précisé leurs calculs, ont estimé qu'il s'accordait mieux à leurs données. Ce modèle a joui de la faveur populaire sous le nom de modèle du H. Fuller, Doctor et Webb⁽⁷⁴⁾ en ont envisagé plusieurs variantes sans parvenir à un choix définitif.

Dans ce travail, plusieurs modèles 4-4 ont été construits. Ils donnent, dans leurs courbes d'intensité théorique, une épaule au même endroit que les modèles 6-2, un peu moins accentuée, mais tout à fait nette.

J'ai eu également l'occasion d'effectuer des relevés de coordonnées sur un modèle 4-4 construit principalement par J.-P. Waller et R. Taglang. Dans tous les modèles 4-4, l'esthétique pêche d'un côté ou de l'autre: les segments qui raccordent les régions ordonnées n'ont pas de configuration satisfaisante. Ou bien, il est possible d'empiler convenablement leurs bases, mais alors la molécule est très ouverte, et il lui correspond une épaule très prononcée, ou bien l'arrangement des bras est choisi de telle manière qu'il impose une structure compacte, et alors les segments de raccordement s'enchevêtrent d'une manière peu plausible.

c) Modèles à trois filaments.

Plus précisément, il s'agit de modèles dérivés de l'arrangement en trèfle, et contenant des régions à trois filaments. Leur raison d'être sera indiquée au chapitre suivant. Les arrangements 5-3 comme les arrangements 3-2-3 donnent une courbe d'intensité théorique sans épaule, en assez bon accord avec les expériences de diffusion centrale.

c) Modèles 2-4-2.

La première idée de modèle où trois régions hélicoïdales sont empilées revient sans doute à Crick (entretien privé). C'est ainsi qu'il conciliait l'idée (communiquée par V. Luzzati) que le tARN est long et mince, avec l'arrangement en trèfle. Quand nous avons proposé le modèle 4-2-2, plusieurs chercheurs ont repris l'idée de Crick en la précisant⁽⁴³⁻⁴⁶⁾.

Dans cette gamme, deux modèles ont été construits. Le premier, avec Favre et Yaniv, sur une suggestion de Crick. Nous avons pu vérifier que la liaison ThioU-C

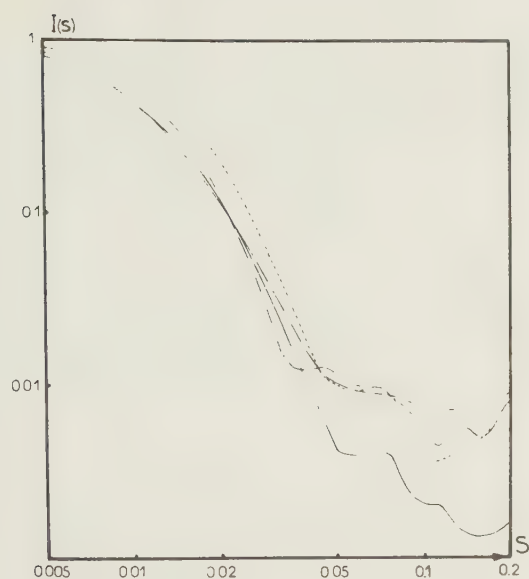


Fig. 3-g. Modèles 2-4-2 comparés au tARN^{Val} chargé (---) et non chargé (—). Le modèle 2-4-2 de Danchin (----) fait 20 Å de Rg. Un autre, construit avec Favre et Yaniv (-.-) fait 22,5 Å.

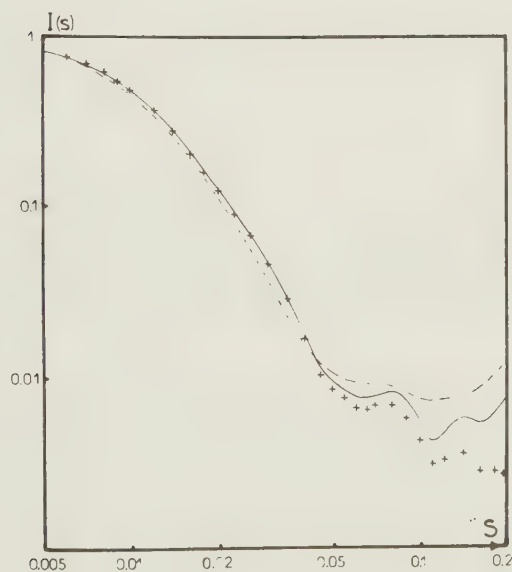


Fig. 3-h. Un modèle 3-2-3 (trait continu) comparé au phénoxy-acétyl-tARN^{Val} (traitillé). Le + correspond à l'intensité théorique de la courbe avant la correction par la fonction intensité du cube de 3 Å de côté.

y est réalisable, mais au prix d'une complexité plus grande que celle du modèle 4-2-2 .

Le second a été construit avec Danchin, et tentait de combiner les propriétés topologiques du modèle de Connors avec certains appariements supplémentaires suggérés par Cramer ou Levitt.

Le point commun à tous ces modèles est qu'ils sont extrêmement compacts. Toutes choses égales par ailleurs, il y a plus de 4 Å de différence entre les Rg du modèle de Danchin, et celui du modèle 4-2-2 proposé au chapitre 7 . Le Rg du modèle de Levitt est compris entre 22,5 et 23 Å ; celui du modèle de Danchin est proche de 20 Å .

Les modèles 2-4-2 ne présentent pas d'épaule à $s=0.03$, mais leur rayon de gyration ,trop faible, suggère que la molécule est plus asymétrique, ou plus aérée que dans ces modèles.

3) SIMULATION DES MODELES MOLECULAIRES.

En principe, une fois que la construction d'un modèle moléculaire de tARN est achevée, il suffit de mesurer les coordonnées et de les confier à l'ordinateur qui se charge d'en déduire la fonction de Patterson. Négligeant les atomes d'hydrogène, il faudrait évaluer 5000 coordonnées dont une grande partie se rapportent à des atomes peu accessibles, et disposer de systèmes de visée élaborés, par exemple de lunettes montées sur pieds et barres coulissantes, pour atteindre une précision meilleure que le quart de centimètre, qui représente 0,8 Å sur les modèles C.P.K. On voit tout de suite qu'il faut trouver des raccourcis.

Le procédé employé par Lake et Beeman consiste à photographier la molécule sous deux angles différents, et à évaluer sur les clichés, les coordonnées des atomes par paquets : un point pour le groupement phosphate, un pour le sucre, et un pour la base.

Dans ce travail, nous avons procédé de manière plus souple et plus rigoureuse. Trois points non alignés suffisent pour définir la position d'un solide. C'est à partir de cette propriété que le programme de simulation des modèles a été écrit. Soit un solide auquel sont liés trois points, A,B,C, et qu'on veut placer de telle manière que ces points occupent les positions respectives P,Q et R .Il existe plusieurs manières de combiner translations et rotations pour y arriver. Dans la procédure adoptée, deux rotations font suite à une translation :

-une translation amène A en P, ce qui conduit B en B' et C en C' .

-les segemnts PB' et PQ sont amenés en coïncidence par une rotation autour d'un axe qui leur est perpendiculaire ; B' vient alors en B'' et C' en C'' .

-les plans $PB''C$ et PQR sont amenés en coïncidence par une rotation autour de PQ .

Naturellement, dans le calcul, les deux rotations, qui se font autour de deux axes concourants, sont combinées en une seule matrice. La puissance de la méthode tient en ceci qu'elle permet de travailler avec des coordonnées mesurées approximatives mais judicieuses, et de corriger les imprécisions en cours de route.

Supposons par exemple qu'on construise avec des modèles C.P.K. deux tronçons d'hélice à deux filaments qu'on voudrait placer verticalement côte à côte dans la configuration de l'ARN IO . On peut choisir trois points initiaux pour fixer la position du premier tronçon, par exemple :

- A, point milieu de la liaison sucre-base sur la base + du premier plateau.
- B, point milieu de la liaison sucre-base sur la base - du premier plateau.
- C, intersection de l'axe de l'hélice avec le plan de la $n^{ième}$ paire de bases.

Il faudra faire coïncider les points théoriques A,B,C, définis par les données cristallographiques de l'ARN IO, avec les points P,Q,R, repérés sur le modèle. On pourra pour le second tronçon choisir des points D,E,F qui sont les mêmes que A,B,C, dans le même ordre ou bien dans un ordre différent. Supposons de plus que dans le modèle construit, il y a, près des points P et P' une région d'encombrement stérique, où la géométrie doit être définie avec soin. Alors, on prendra P et P' comme premiers points respectifs. Comme derniers points, on pourra choisir des points du type C, intersections de l'axe de l'hélice avec le plan des bases. Soit Oz la direction de cet axe; puisqu'on a supposé les tronçons verticaux, on déduira z_C de z_A et z_B ; ou plus exactement, on apportera de petites corrections aux valeurs mesurées. Sur x_C et y_C , on peut commettre des erreurs tout-à-fait grossières, sans que cela porte à conséquence, car la seule chose qui importe est que ces coordonnées définissent un point qui soit du bon côté par rapport à la droite AB . Et si on se trompe, on aura construit une hélice qui descend au lieu de monter, et l'erreur sera facile à déceler et corriger. Si on veut simuler un tronçon incliné, on aura intérêt à prendre le troisième point sur un plateau éloigné du plateau des deux premiers.

Les régions en simple ou en double-hélice utilisées dans les simulations ont toutes été empruntées aux données cristallographiques : ADN de forme A, ou B, ARN IO ou II . Pour les régions en triple-hélice, on a soit adjoint aux deux filaments de l'ARN IO un troisième filament par rotation autour de l'axe, soit utilisé un modèle construit à la main, puis standardisé. Dans cette dernière

opération, on a considéré que la base était un solide, et le groupement sucre-phosphate un autre solide; on a ainsi ajusté par la méthode trois points-trois points ces deux solides, prélevés sur l'ARN IO, à chaque filament du modèle construit.

Pour la simulation des boucles coiffant les régions hélicoïdales, on faisait coïncider le dernier plateau de bases du tronçon hélicoïdal avec le premier plateau (virtuel, mais construit) de la boucle.

En fait, les opérations de simulation de modèles pour le tARN n'étaient pas conduites d'emblée avec les coordonnées de tous les atomes. Au lieu de cela, on employait un plateau de bases fictif, appelé "squelette", et composé des points suivants :

- 1-intersection de l'axe de l'hélice avec le plan du premier plateau.
- 2-intersection de l'axe de l'hélice avec le plan du cinquième plateau.
- 3,4,5 -phosphates 3' sur la base +, la base-, et la troisième base.
- 6,7,8-point milieu de la liaison sucre-base sur les trois bases précédentes.
- 9,10 -phosphate 3' sur la base + et la base- du plateau suivant.

De la sorte, on peut savoir au premier coup d'oeil si l'hélice va dans la bonne direction (le point 1 voyage en ligne droite) et si elle est bien construite (le point 9 d'un plateau doit coïncider avec le point 3 du plateau suivant). Le dernier plateau d'un tronçon hélicoïdal fournit généralement les coordonnées des points 1, 9 et 10 qui servent à positionner la boucle qui le coiffe.

Pour les régions non standardisées (nucléotides de raccord; parfois la boucle DiHU), les positions sont également simulées par la méthode trois points-trois points, en considérant que le nucléotide tout entier est dans la configuration de l'ARN IO ou II. Cette hypothèse a également été faite pour les boucles. Les points choisis pour fixer le solide sont : le point milieu de la liaison sucre-base, le phosphate 3' et le phosphate 5'. Le résultat est évidemment approximatif, mais on a l'assurance que les phosphates sont bien placés, et que la base est bien orientée par rapport à la surface de la molécule.

En travaillant de cette manière, et quand on dispose de tous les composants (régions hélicoïdales, boucles), la simulation d'un modèle est rapide : il faut compter deux ou trois heures pour mesurer les coordonnées nécessaires, et une demi-journée de travail en ordinateur (dont quelques secondes de calcul en temps réel) .

4. INTENSITES THEORIQUES CORRESPONDANT A DIFFERENTS MODELES.

Les intensités théoriques pour différents modèles d'oligonucléotides sont

données sur les Fig. 3-a à 3-d. La normation est telle qu'à l'origine, $I(0)$ est l'équivalent d'une "masse apparente". Plus précisément, $I(0)$ est égal, sur ces figures, au produit du nombre d'électrons apparents par 1,966.

Les intensités théoriques pour différents modèles de tARN sont données sur les Fig. 3-e à 3-h. La normation est telle que $I(0) = I$. L'intensité théorique correspondant au modèle 4-2-2 est donnée au chapitre 7.

- CHAPITRE 4 -

PROPRIETES DES REPRESENTATIONS D'ACIDES NUCLEIQUES

I - INTRODUCTION :

Quand on observe la représentation d'une protéine de structure connue, comme le lysozyme, et qu'on essaie par la pensée de suivre l'enroulement de la chaîne polypeptidique dans ses méandres, on se prend à douter qu'il soit possible de comprendre d'une manière simple la conformation d'une protéine, et de prédire ses propriétés fonctionnelles quand on ne connaît que sa structure primaire.

Le chimiste organicien qui étudie une molécule compliquée peut se satisfaire, pour l'imaginer, d'un dessin sur le papier. Par contre, le biologiste qui étudie les macromolécules est confronté à des formes tellement complexes qu'il a du mal à les reconnaître. Vue sous un angle déterminé, la molécule d'hémoglobine est aisément reconnaissable. Mais je doute que même les spécialistes réussissent à s'en faire une représentation aussi familière que celle d'un objet domestique complexe, ou celle d'une macromolécule régulière comme la double-hélice d'ADN. La collection de structures cristallographiquement déterminées de macromolécules s'enrichit sans cesse, mais leur "assimilation" ne progresse guère.

Notre manière de comprendre et de voir les acides nucléiques est enfermée dans un carcan, celui des concepts tirés des structures cristallographiques d'ADN ou d'ARN en double-hélice. Bientôt viendront les structures beaucoup moins régulières du tARN, de l'ARN 5s ou 6s. Un jour peut-être, on voudra connaître les détails de l'enroulement de l'ARN dans les ribosomes. Cependant, la situation se présente de manière beaucoup plus favorable que dans le cas des protéines. La raison en est qu'on connaît déjà des principes de construction extrêmement simples, et dont la validité semble bien établie. Par exemple, deux régions de la molécule peuvent former des associations qui engagent des ensembles de nucléotides consécutifs.

L'étude entreprise dans ce chapitre a pour objectif d'affiner la représentation des acides nucléiques afin de s'en faire une image utilisable. Elle cherche à dégager des règles qui permettent, connaissant la structure primaire et un arrangement secondaire, de prévoir les différentes classes d'arrangements tridimensionnels qui y satisfont.

La portée de cette étude est éminemment pratique, dans le cadre d'une construction de modèles de tARN. En effet, une dizaine de modèles ont été proposés, et il importe de savoir avec quel degré d'élasticité les propriétés du modèle sont liées aux hypothèses de départ. Peut-on modifier une partie du modèle, changer tel ou tel paramètre sans trop enfreindre d'autres règles données au départ ? La nécessité de l'étude entreprise ici apparaît également quand on examine certains arrangements proposés pour l'ARN 5s. Par exemple, celui qu'a suggéré Cantor⁽⁷⁶⁾ est tellement complexe qu'on ignore s'il est possible d'en faire un modèle tridimensionnel qui obéisse à des contraintes physico-chimiques raisonnables.

Quand on construit un modèle, il y a dans chaque liaison que l'on fige, une hypothèse implicite ou explicite ; et comme il est impossible d'étudier systématiquement toutes les possibilités, il importe de connaître la limite de validité des résultats obtenus et encore mieux, de discerner les hypothèses auxquelles on fait appel. Quatre classes de propriétés seront ici distinguées :

1) - Propriétés topologiques.

Ce sont celles dans lesquelles l'acide nucléique est considéré comme un fil : on ne se soucie ni de la séquence des bases ou de leur orientation, ni de la configuration de la chaîne des sucres-phosphates. Le fil peut être ouvert ou fermé et les associations des différentes régions de la molécule peuvent donner lieu à des diagrammes variés. Enfin, signalons que les problèmes liés à la circularité dans l'ADN ont fait l'objet de nombreuses études, et que celle de la présence éventuelle de noeuds dans certains acides nucléiques mérite d'être soulevée.

2) - Propriétés d'orientation.

Ce sont celles pour lesquelles on considère l'acide nucléique comme une chaîne orientée d'une extrémité 5' à une extrémité 3', et à laquelle sont attachées des bases orientées entre elles et par rapport à la chaîne. Des problèmes pratiques de construction de modèles peuvent être résolus par une analyse soignée des propriétés d'orientation des chaînes libres ou associées. Un exemple malheureux de construction de modèles à l'envers a été cité par Crick⁽⁷⁵⁾.

3) - Propriétés métriques.

Ici, la géométrie détaillée des acides nucléiques est prise en considération, et en particulier les angles et distances de liaison. Les problèmes pratiques trouvant leur place dans cette étude sont, par exemple : sous quelles conditions une paire G.U est équivalente à une paire U.G ; quelles sont toutes les structures hélicoïdales régulières,

parallèles ou anti-parallèles qu'il est possible de former avec des segments d'acides nucléiques ?

4) - Propriétés de séquence.

On peut se demander si certaines séquences sont caractéristiques des boucles, d'autres se trouvant de préférence dans les segments ordonnés. Une étude intelligente des différentes séquences publiées, soit de tARN, soit de fragments d'ARN messenger devrait fournir des informations précieuses.

On trouvera ici une étude détaillée des deux premières classes de propriétés.

II - PROPRIETES DE TOPOLOGIE^(75b)

Les acides nucléiques peuvent être décrits comme des chaînes ou des filaments parce que leur nature chimique est celle de polymères linéaires. La chaîne peut être ouverte, et dans ce cas, les biochimistes distinguent deux types d'extrémités que nous n'avons pas besoin de caractériser ici. Parfois l'acide nucléique est en chaîne fermée : c'est le cas de l'ADN dans le chromosome d'E. coli tel qu'il est vu par autoradiographie, de l'ADN mitochondrial, de certains acides nucléiques viraux vus par microscopie électronique. Il a été démontré dans certains cas que l'apparence annulaire correspond effectivement à la présence de chaînes fermées, ininterrompues. Nous dirons dans ce cas que l'ADN ou l'ARN est circulaire.

Que la chaîne soit ouverte ou fermée, il est possible de définir un critère d'intériorité pour un quelconque de ses segments. Soient a, b et c trois positions (trois nucléotides) sur le filament. Définissons $D(a,b)$, la distance entre a et b égale au nombre de nucléotides qui sépare a de b, plus un. En outre, $D(a,a) = 0$.

Dans le cas d'une chaîne fermée, cette définition conduit à deux valeurs, et on la précise en choisissant la plus petite. Il existe encore une ambiguïté quand la chaîne possède un nombre pair de nucléotides : dans certains cas, les deux valeurs trouvées peuvent être égales. On supprime l'ambiguïté sur le chemin, en rajoutant un nucléotide "blanc".

Dans tous les cas, si :

$$D(a,b) = D(a,c) + D(c,b),$$

nous dirons que c est à l'intérieur du segment ab. Le segment ab est la réunion de tous les nucléotides c obéissant à l'égalité ci-dessus, y compris a et b. La notion de distance

n'est pas employée ici pour représenter une distance physique (car $D(a,b)$ peut être grand pour deux nucléotides a et b qui se touchent) mais pour mieux définir les relations d'inclusion. Dans le cas d'une chaîne fermée, nous prendrons bien garde à ne pas donner à l'intérieur et à l'extérieur un sens absolu, mais à n'utiliser l'un que par rapport à l'autre.

En solution, une chaîne comme celle du tARN, a une structure secondaire et perd en quelque sorte une partie de sa linéarité du fait que les liaisons hydrogène se forment entre les bases. Mais celles-ci sont beaucoup plus labiles que les liaisons covalentes et se manifestent d'une autre manière. Nous considérons que la linéarité de la chaîne n'est pas affectée par les liaisons hydrogène, et nous dirons que deux régions entre lesquelles s'échangent des liaisons hydrogène sont associées.

a - Contiguité et parallélisme dans les associations.

Considérons maintenant les diverses figures topologiques auxquelles peuvent donner lieu les associations. Soient $A_1, A_2, \dots$ An les segments adjacents correspondant aux nucléotides successifs de la chaîne ; la réunion d'un nombre quelconque de segments adjacents est aussi un segment.

Si la chaîne est fermée, A_n est adjacent à A_1 .

Imaginons que A_p est associé à A_q . Des deux segments A_{p-1} et A_{p+1} , l'un est intérieur au segment d'extrémités A_p et A_q (que nous appellerons A_{pq}) et l'autre lui est extérieur. Nous pouvons distinguer différents cas :

- C1) A_{q-1} est associé à A_{p+1} ou A_{q+1} est associé à A_{p-1} ou les deux.
- C2) A_{q-1} est associé à A_{p-1} ou A_{q+1} est associé à A_{p+1} ou les deux.
- C3) Il existe un A_u à l'intérieur de A_{pq} associé à un A_v extérieur.
- C4) Il n'existe pas d' A_u intérieur à A_{pq} associé à un A_v extérieur.

Nous considérons que C1 définit une association anti-parallèle et que C2 définit une association parallèle. Cette définition s'applique pour l'instant aux associations intrachaines seulement, mais elle remplace avantageusement la définition classique qui suppose un ordre cristallographique dans les associations semblable à celui de l'ADN, et dont on est rarement assuré. Une association ne comprenant qu'une seule paire de bases ne répond ni à C1 ni à C2 ; ceci n'est guère gênant.

Avec les définitions proposées ici, les associations du "trèfle" sont toutes anti-parallèles. Par contre, certaines associations supplémentaires postulées par Cramer (AG-TV) ou par Levitt⁽⁴³⁾ sont parallèles. On trouvera des associations composées d'une seule paire de bases dans divers arrangements relatifs à l'ARN 5s. Mais tous les arrange-

Fig. 4-a . Associations antiparallèles (a1) ou parallèles (a2) dans une chaîne isolée. La définition est relative à l'ordre de succession des nucléotides dans les régions associées, mais non à la géométrie locale.

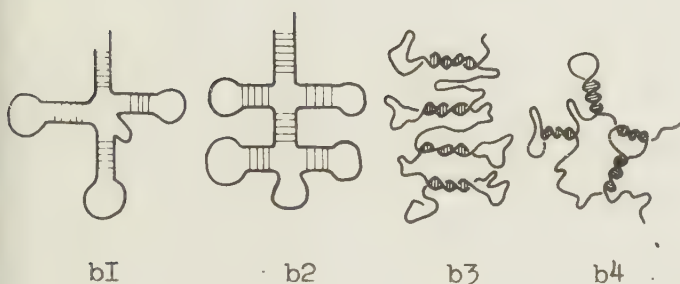
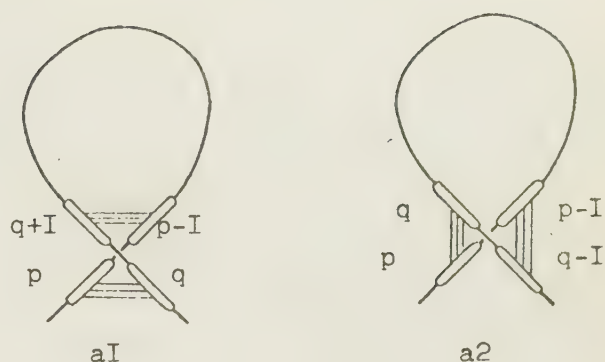


Fig. 4-b . Le diagramme du trèfle pour le tARN (b1) et un diagramme similaire pour l'ARN 5s proposé dans la réf. satisfont au principe de contiguité, de même que le schéma proposé par Spirin (ref.) pour l'ARN ribosomal à force ionique élevée (b3). Par contre, le schéma proposé par Spirin pour l'ARN ribosomal à faible force ionique, n'y satisfait pas.

Fig. 4-c .Après reformulation du principe de contiguité, la situation demeure inchangée pour le diagramme c1 .Par contre, c2 qui contient des associations parallèles satisfait maintenant à ce principe.

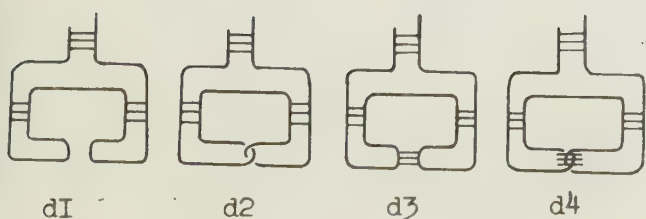
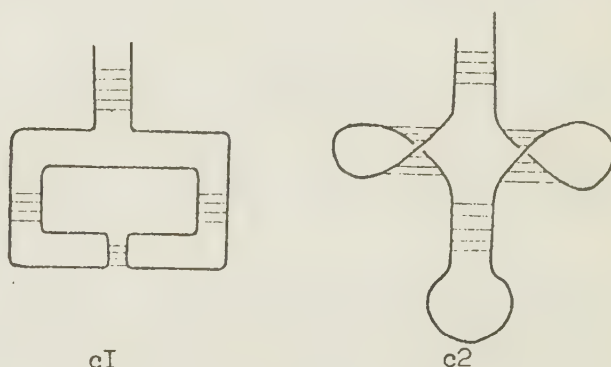


Fig. 4-d .Illustration du lien qui peut exister entre les diagrammes d'associations, et les situations topologiques qui leur correspondent de manière non univoque (voir texte).

ments (76-79) proposés pour cette molécule paraissent peu convaincants, et on peut se demander s'il ne conviendrait pas d'examiner les possibilités d'associations parallèles selon des règles d'appariements qui ne seraient plus, évidemment, celles de complémentarité, rigoureuse ou approchée.

Quand on sait que deux segments sont associés, mais qu'on ignore le détail des associations entre les nucléotides qui les constituent, on ne peut pas dire si l'association est parallèle ou anti-parallèle. Ceci est mis en relief dans la similitude des dessins de la Fig. 4-a. La difficulté rencontrée ici est la même que celle relative à la définition d'une paire de bases isolée.

Considérons maintenant les critères C3 et C4.

C4 est satisfait dans l'arrangement en trèfle, dans ceux proposés pour l'ARN 5s sauf deux, dans celui suggéré pour l'ARN 6s. Il est également respecté dans le schéma proposé par Spirin⁽⁸⁰⁾ pour l'ARN ribosomal à force ionique élevée mais pas dans celui qui correspond aux faibles forces ioniques (Fig. 4-b).

Par ailleurs, C2 et C4 sont incompatibles. Or, si nous regardons la molécule représentée dans la Fig. 4-c₁ nous voyons qu'elle satisfait à C2, mais qu'elle satisferait aussi à C4 pourvu que, lorsque nous parlons de Ap, Aq..., il ne s'agisse plus des nucléotides adjacents membres d'une même association. Nous appelons dorénavant "région", une telle réunion des segments adjacents membres d'une même association, et nous pouvons reformuler C3 et C4 :

C3) Si Ap et Aq sont des régions associées, il existe une région Au intérieure à Apq, associée à une région Av, extérieure à Apq.

C4) Si Ap et Aq sont deux régions associées, il n'existe pas de région Au intérieure à Apq, associée à une région Av extérieure à Apq.

Physiquement, nous avons perdu en résolution ; nous avons cessé de regarder le détail des arrangements à l'intérieur de l'association, et la propriété qui fait l'objet de C1 et de C2 (le parallélisme) devient indépendante de celle qui correspond à C3 et C4. C2 et C4 ne sont plus contradictoires. Nous dirons qu'une molécule dont les régions associées satisfont à C4, obéit au principe de contiguité.

Avons-nous des raisons de penser que les molécules d'ARN suivront en général le principe de contiguité ? Fresco, Alberts et Doty⁽⁸¹⁾ ont même proposé que celles-ci se replieraient d'une manière encore plus restrictive que C4, les associations se formant préférentiellement entre régions consécutives. Nous tenterons de montrer plus loin que certaines précautions sont à prendre quand on propose une structure qui désobéit à C4.

Telles qu'elles ont été définies, les notions de parallélisme et de contiguité correspondent à des idées intuitives et générales, comme l'attestent les propriétés suivantes :

A) - Propriété de réciprocité.

Considérons deux associations dont l'une relie Ap et Aq, et l'autre lie un Au intérieur à Apq, à un Av qui lui est extérieur : la molécule ne suit pas le principe de contiguité. Mais alors, si on considère l'association de Au avec Av, il est clair que Ap et Aq ne sont pas simultanément à l'intérieur ou à l'extérieur de Auv.

B) - Invariance dans les opérations d'ouverture ou de clôture des chaînes.

Les propriétés de parallélisme et de contiguité des associations intra-chaînes ne changent pas lors de la fermeture de la chaîne si elle était ouverte, ou de son ouverture, si elle était fermée.

C) - Invariance lors du raccordement de chaînes ouvertes.

Les propriétés de contiguité et de parallélisme ne changent pas quand on raccorde bout-à-bout des chaînes ouvertes.

D) - Extension des définitions aux associations de plusieurs chaînes.

Considérons maintenant le cas de deux chaînes ouvertes contenant des associations internes, et des associations liant une région de la première à une région de la seconde. Pour définir la contiguité et le parallélisme dans ce cas, transformons les deux chaînes en une chaîne unique en les joignant bout-à-bout. En combinant les propriétés précédemment énoncées, on voit que la contiguité ne dépend pas du choix des extrémités qui ont été connectées. Mais le parallélisme en dépend, et il est clair que les quatre manières de connecter les extrémités sont équivalentes par paires, du fait de l'invariance du parallélisme dans les opérations de clôture ou d'ouverture d'une chaîne isolée.

Par conséquent, une définition complète du parallélisme requiert la distinction des deux extrémités, et relève des propriétés d'orientation. Ceci devient encore plus évident quand on étudie les associations de chaînes fermées.

E) - Formation séquentielle de la structure secondaire.

Comparons les Fig. 4-d1, 4-d2, 4-d3 et 4-d4. d1 et d2 satisfont au principe de contiguité, mais pas d3 et d4. Cependant, les membres de chacune de ces deux paires diffèrent dans leur statut topologique. Pour une molécule donnée, on peut considérer que la situation d1 est moins probable

que d2. En effet, si on considère qu'il existe une certaine souplesse dans la structure tertiaire, la forme d2 est restrictive : elle exige que les deux boucles qui s'interpénètrent occupaient des positions relatives très définies avant la formation d'au moins deux des associations représentées. Par conséquent, si les deux formes sont possibles, il est raisonnable d'estimer que la première prédominera. En revanche, il paraît difficile de se prononcer a priori sur les deux formes suivantes d3 et d4. Les interactions figurées au niveau des boucles contribuent effectivement à l'énergie de formation de la structure. Par conséquent, chaque fois que des associations non contigües sont postulées, il faut s'inquiéter de la séquence de formation des associations et examiner la possibilité de complications topologiques du type d4 qui seront définies plus loin de manière plus précise.

b - Circularité.

Si une chaîne est fermée, elle peut être ou ne pas être topologiquement équivalente à un cercle. Dans le premier cas, il existe une manière de lui appliquer une déformation continue qui la transforme en un cercle ; dans le second cas (Fig. 4-e1) elle forme un noeud et il n'existe pas de déformation continue qui puisse la transformer en cercle ; une telle transformation exige une coupure de la chaîne et une soudure ultérieure des extrémités.

Les associations de deux chaînes circulaires peuvent à leur tour former diverses figures topologiques. En 4-e2, il est possible de séparer complètement les deux chaînes sans coupure préalable, tandis qu'en 4-e3 et 4-e4, une coupure est nécessaire. Nous dirons, avec Thomas et MacHattie⁽⁸²⁾ que les cercles en 4-e3 et 4-e4 présentent un lien topologique d'ordre 1 et 4 respectivement. Des lignes fermées continues ont été observées dans les autoradiographies d'ADN d'E.coli⁽⁸³⁾ et dans les clichés de microscopie électronique d'ADN pour de nombreux phages, depuis les travaux de Ris et Chandler⁽⁸⁴⁾ sur l'ADN du phage lambda, et ceux de Kleinschmidt, Burton et Sinsheimer⁽⁸⁵⁾ sur la forme répliquative du phage ϕX .

Cependant, ces formes annulaires ne résultent pas toujours de l'association de chaînes fermées. Par des méthodes variées, il a été montré que l'ADN du phage λ est une association de deux chaînes du type montré sur la Fig. 4-f1. Par ailleurs, des chaînes topologiquement liées d'acides nucléiques devraient avoir des propriétés remarquables, différentes de celles de duplex linéaires ; ces propriétés sont exposées dans des articles de revue^(82, 85). Par conséquent, dans de nombreux cas, il a été possible d'identifier des molécules d'ADN circulaire topologiquement liées ; de plus, l'ordre de la liaison topo-

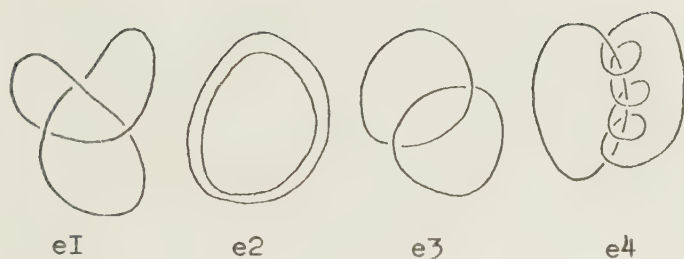


Fig. 4-e . Diagrammes formés par des chaînes fermées. En e1, la chaîne n'est pas équivalente à un cercle. En e2, les deux chaînes "circulaires" sont topologiquement indépendantes, tandis qu'elles présentent un lien topologique d'ordre 1 en e3 et d'ordre 4 en e4 .

Fig. 4-f. Diverses figures topologiques peuvent, à faible résolution, présenter la même apparence annulaire que l'ADN circulaire . Les diagrammes f1, f2, f3 contiennent deux chaînes ouvertes . f4 et f5 contiennent une chaîne fermée unique , tandis qu'en f6 , la chaîne est ouverte. Un système enzymatique capable de transformer un duplex linéaire d'ADN en une structure de type f4 est décrit dans la ref. 87

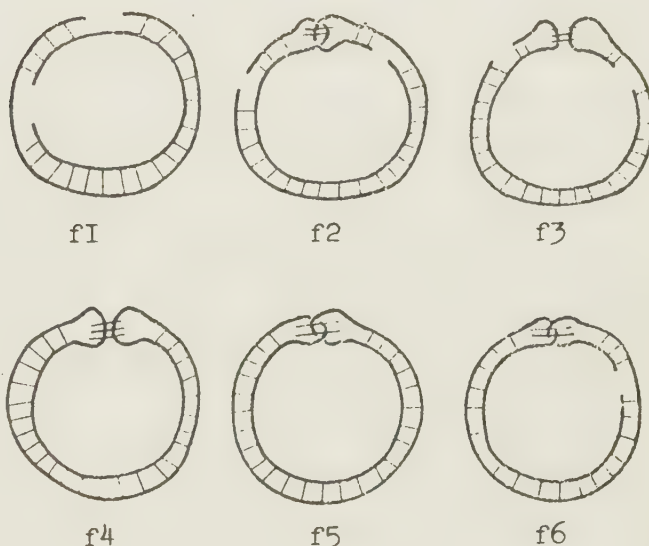


Fig. 4-g. L'apparence d'un court segment d'ARN en double-chaîne, circulaire, contenant 21 paires de bases , dessinée d'après un modèle C.P.K.

gique a pu être estimé, et trouvé voisin du nombre de tours d'hélice que ferait le duplex s'il était linéaire. En d'autres termes, un ADN circulaire a localement une configuration voisine de celle qu'il a sous forme linéaire ; ceci est plus surprenant qu'il paraît à première vue, car on comprend mal dans ces conditions comment peut se faire la séparation des brins en vue de la réplication. Dans la bactérie, la réplication de l'ADN est semi-conservative et par conséquent, elle implique à un moment ou à un autre une séparation complète des brins. D'autre part, Cairns⁽⁸³⁾ a montré que durant la réplication, les chaînes conservent une forme annulaire, même quand la séparation s'est produite sur une large fraction du génôme. Comme il est peu vraisemblable que les contraintes s'accumulent dans la région répliquée en dernier, Cairns a postulé l'existence d'un système qui coupe le cercle en un point et laisse pivoter les deux extrémités pendant que la réplication se poursuit. Cette hypothèse n'a pu encore être vérifiée, et il est loisible d'imaginer des statuts topologiques variés pour des ADN d'apparence annulaire (Fig. 4-f). Il semble cependant que les enzymes intervenant dans la réplication de l'ADN bactérien sont variés, et cela suggère une certaine complexité dans le processus.

Le mécanisme de réplication des acides nucléiques viraux^(88, 89) pourrait différer du mécanisme bactérien, et le modèle du "cercle qui roule" proposé par Gilbert et Dressler⁽⁹⁰⁾ pourrait avoir une validité étendue. Dans ce modèle, une chaîne est ouverte, l'autre fermée durant la réplication.

En raison de la complexité des processus qui peuvent rendre compte de la réplication des acides nucléiques circulaires, on peut se demander quelle est la raison d'être de la circularité observée. Des avantages sélectifs divers ont été invoqués^(91, 92), mais n'entraînent pas la conviction.

Cette étude topologique ne suggère aucune explication simple. Toutefois, l'existence des acides nucléiques circulaires pourrait être due à des propriétés simples et fondamentales qui nous échappent. Il serait intéressant d'étudier la variation des propriétés physiques des ADN circulaires avec la longueur des chaînes. Des propriétés nouvelles pourraient apparaître avec les chaînes très courtes qui méritent d'être considérées dans le contexte du problème des origines de la vie. Une étude sommaire de construction de modèles montre qu'il est difficile de construire une double-chaîne d'acide nucléique contenant moins de 20 paires de bases, si on veut conserver leur empilement. Un modèle contenant 21 paires de bases est représenté sur la Fig. 4-g. L'ordre de la liaison topologique est de deux.

c - Les noeuds.

Un noeud est une forme compliquée que le sens commun identifie aisément, mais

qu'il est difficile de définir. En particulier, il n'existe aucun critère analytique général permettant de dire si une courbe fermée forme un noeud ; et s'il existe déjà une certaine classification des noeuds, celle-ci n'est pas exhaustive et on ignore même parfois si tel noeud n'est vraiment pas réductible topologiquement à tel autre.

En 4-h3, est représentée l'idée intuitive du noeud le plus simple qu'on puisse concevoir. Cependant, si on compare 4-h1 avec 4-h2 et 4-h3, on voit que cette conception intuitive n'est pas purement topologique. Si nous raccourcissons les extrémités de la chaîne, le noeud disparaît. Inversement, il est possible de transformer le filament de 4-h1 en filament porteur de noeud, en tirant sur les extrémités. C'est d'ailleurs ainsi qu'en pratique, on manipule le noeud. Plus les extrémités sont longues, et plus la forme 4-h2 ou 4-h3 prend allure de noeud. Prolongeons les deux extrémités indéfiniment sur la même droite : de cette façon, nous transformons une chaîne ouverte en chaîne fermée (sur une droite, il n'y a qu'un seul point à l'infini), et on aboutit à la figure 4-h4 qui représente un noeud topologique vrai. Nous nous inspirerons de cette démarche pour la discussion qui suit : nous regarderons des structures compliquées, nous leur appliquerons certaines transformations, et nous verrons si le résultat est un noeud topologique.

1) - Noeuds topologiques vrais.

Ceux-ci ne peuvent se trouver que dans les chaînes fermées. Vinograd et al (26) se sont demandés si certains ADN circulaires formaient des noeuds. Jusqu'à présent, à ma connaissance, un pareil cas n'a jamais été signalé par les microscopistes. Il paraît extrêmement difficile de tenter d'évaluer la probabilité pour qu'une chaîne fermée engendrée par un processus aléatoire bien défini, fasse un noeud. On peut cependant faire quelques remarques :

- I - Définissons un type très particulier de divagation (random walk) dans un réseau cubique. Chaque pas est de longueur unité le long d'une des trois directions d'un trièdre orthonormé, et on interdit au chemin de passer deux fois par un même point, autre que l'origine, qu'il rejoint au dernier pas. Le nombre minimum de pas requis pour faire un noeud est de 24, et la probabilité pour un tel chemin aléatoire, fermé, de 24 pas, de faire un noeud semble très faible. Delbruück à propos des mêmes questions, a été conduit à définir une divagation dans un réseau cubique un peu différent de celui propose ici⁽⁹³⁾. Le nombre minimum de pas est alors de 36.
- II - Un ADN circulaire, long et flexible, visible en microscopie électronique, a été obtenu par Thomas et MacHattie⁽⁹⁴⁾ ; le matériel de départ était de l'ADN du phage T2 linéarisé par coupure. Mais la coupure s'étant produite au

hasard sur la chaîne, la séquence est définie à une permutation circulaire près. Par dénaturation et renaturation thermique de ce matériel, on obtient un ADN annulaire beaucoup plus long que le matériel d'origine, contenant de nombreuses "coupures" sur chaque chaîne. Des noeuds pourraient y être recherchés mais l'intérêt biologique de cette étude serait plutôt lointain.

2) - Noeuds par circularisation.

Dans l'ARN de transfert, comme dans les ARN 5s et 6s, les deux extrémités sont associées. Quand on les joint par une liaison covalente fictive, en ligne droite, on peut demander si la chaîne ainsi transformée présente un noeud. La réponse est positive dans le cas du schéma de la Fig. 4-11 représentant selon Spirin⁽⁸⁰⁾ l'ARN ribosomal à faible force ionique. La molécule trouvée dans cet état est considérée comme "désordonnée". Or, ce type de configuration peut au contraire sembler assez peu représentatif d'un état désordonné, même s'il est proposé que cette configuration n'est qu'un échantillon d'une grande variété de configurations similaires. Leur formation peut être envisagée de deux manières différentes :

- I - Les associations qui donnent naissance aux boucles se produisent après que celles-ci se soient associées à d'autres régions de la chaîne. Donc, plusieurs couples de régions complémentaires éloignées se reconnaissent avant plusieurs couples de régions complémentaires voisines.
- II - L'ordre est inverse du précédent, mais une extrémité doit se frayer un chemin à travers les boucles.

Il est vraisemblable que la raison pour laquelle Spirin fait passer un filament à travers une boucle tient au fait que si les associations sont, comme on le suppose généralement, hélicoïdales, un filament s'enroule autour de l'autre. Dès que l'association dépasse une certaine longueur, le phénomène est presque inévitable.

Si on essaie de construire un modèle tridimensionnel à partir de l'arrangement proposé par Cantor pour l'ARN 5s, plusieurs situations topologiques sont possibles. La règle du jeu étant qu'un filament peut traverser (ou non) une boucle s'il forme avec elle une association, il existe des cas où il en résulte un noeud par circularisation. Par contre, la molécule représentée sur le Fig. 4-13 n'est pas assez complexe pour donner naissance à un noeud par circularisation, malgré la concaténation de deux boucles. Cette figure représente le modèle moléculaire proposé par Levitt⁽⁴³⁾ pour le tARN, mais ne reflète pas exactement la pensée de l'auteur, une erreur matérielle s'étant glissée dans le dessin (Levitt, communication personnelle).

3) - Noeuds métriques.

Quand, en examinant la Fig. 4-13, on dit que deux boucles sont concaténées, il y a une représentation sous-jacente des régions associées par des solides impénétrables (Fig. 4-j). Il est possible de prendre la molécule dans l'arrangement du trèfle, solidifier les quatre associations, puis introduire une cassure dans chaque solide. On obtient ainsi quatre morceaux dans le cas habituel, et trois morceaux seulement dans le cas de la Fig. 4-13. En réalité, quand les associations ont été solidifiées, un choix a été fait, qui ne tenait compte que des associations du trèfle, et non des associations supplémentaires qui peuvent être proposées dans un modèle particulier. Le choix ainsi fait répond à un critère topologique qui le légitime ; il répond à la question générale suivante : soit une molécule contenant des associations non contigües ; quel est le nombre minimum d'associations à supprimer afin que toutes les associations restantes deviennent contigües ? Quelles sont les différentes solutions correspondant à ce minimum ?

Dans le cas Levitt, l'association des boucles est la seule qu'il faille supprimer.

Nous n'abordons pas l'étude générale du problème, qui est un cas d'application de la théorie des graphes. Les données expérimentales ne sont pas assez nombreuses pour justifier pareille étude. Il y a enfin un cas qu'on peut ranger dans la catégorie des noeuds, à condition de trouver encore une définition à ce terme : c'est celui du modèle de la Fig. 4-kl.

4) - Noeuds élastiques.

Ce genre de modèles peut être considéré de la manière suivante : les régions en double-hélice sont des solides, et les boucles sont élastiques. Entre les régions ordonnées, il y a de petites charnières inélastiques. Avec ces conventions, on voit qu'il y a une restriction aux déplacements des régions ordonnées, parce que la région ordonnée DiHU est emmarrée, par sa boucle à l'extrémité acceptrice. La structure se formerait à partir du trèfle ouvert parce que la boucle a une certaine élasticité ; elle peut s'allonger, entourer le bras accepteur, puis se contracter ; le "noeud" se défait aussi difficilement qu'il se fait.

De tous les noeuds passés en revue, celui-ci est incontestablement le plus justifié d'un point de vue physique. Selon Danchin⁽⁹⁵⁾, les curieuses cinétiques de T-jump observées sur le tARN ainsi que les expériences de RPE suggèrent que la formation de la structure secondaire du tARN s'achève par un événement très improbable, qui pourrait être la formation d'un tel noeud.

III - PROPRIETES D'ORIENTATION.

a) - Orientation de la chaîne considérée comme un fil.

Dans les acides nucléiques connus, deux nucléotides successifs sont rattachés par pont phosphorique entre le carbone 3' de l'un et le carbone 5' de l'autre. A chaque extrémité d'une chaîne ouverte, il existe un ribose dont l'un des deux carbones 3' ou 5' n'est pas relié par pont phosphorique à un autre nucléotide. Ce carbone donne son nom (3' ou 5') à l'extrémité. Dans le tARN, il y a ainsi deux extrémités, appelées 3'OH et 5'p. La raison en est que du côté 3', le carbone porte un hydroxyle, et que du côté 5', il porte un groupement phosphate qui lui est resté du nucléotide triphosphate à partir duquel il a été rattaché à la chaîne.

Par convention, on oriente la chaîne de l'extrémité 5' libre vers l'extrémité 3' libre. Si on considère deux nucléotides qui, dans ces conditions, sont successifs, on voit que le sens de progression de la chaîne est du carbone 3' de l'un au carbone 5' du suivant.

Souvent, et cela prête à confusion, on entend parler à propos de l'ADN, d'une chaîne qui monte quand elle est 3'-5' ou qui descend quand elle est 5'-3'. Cela veut généralement dire que dans la chaîne montante, la progression se fait du C3' d'un nucléotide au C5' du nucléotide suivant, donc de l'extrémité 5' libre à l'extrémité 3' libre.

La convention relative aux codons est de les écrire de gauche à droite, de l'extrémité 5' libre à l'extrémité 3' libre. AUG, codon de la méthionine se détaille ainsi : A 3'-5' U 3'-5' G 3'. La même convention ne s'est pas tout-à-fait établie pour les anticodons, mais tend à se généraliser : l'anticodon de la méthionine chez E.coli est donc CAU. Pour passer de l'un à l'autre, il faut un petit effort d'attention car les bases appariées n'occupent pas le même rang, mais on s'y fait rapidement.

L'usage est également établi de représenter les séquences d'acides nucléiques de l'extrémité 5' libre vers l'extrémité 3' libre et celle des protéines, du NH₂ libre vers le COOH libre. De cette façon, les deux catégories de séquences sont orientées dans le sens de la traduction.

L'alternance régulière des carbones 3' et 5' sur la chaîne est due en partie aux conditions enzymatiques de biosynthèse. Les recherches relatives aux polycondensations non enzymatiques de nucléosides diphosphates conduisant fréquemment à des liaisons 2'-5'⁽⁹⁶⁾. Par conséquent, on pourrait avoir à traiter par la suite de chaînes dans lesquelles il n'est pas possible de définir une orientation globale. Par contre, il demeure possible de définir localement l'orientation.

Les oxygènes portés par les carbones 2' et 3' du ribose sont situés d'un même côté par rapport au plan du pentose tandis que le carbone 5' se trouve de l'autre. Nous

établissons une relation d'équivalence entre les positions 2' et 3'. Pour chaque tronçon reliant une paire de nucléotides, on regarde les carbones liés au phosphate intermédiaire. Si aucun d'eux ne porte le numéro 5', on renonce à orienter localement la chaîne. Sinon, on l'oriente dans le sens des n croissants.

Dans les chaînes fermées, il est possible de la même manière de définir une orientation locale. Une chaîne d'ADN obtenue par circularisation d'une chaîne ouverte est évidemment telle que les orientations locales déterminent un sens de parcours unique sur toute la chaîne.

b) - Orientation affectée à une base.

Les atomes d'une purine comme ceux d'une pyrimidine peuvent être numérotés. Adoptons ici la convention utilisée par Michelson⁽⁹⁷⁾. Sur chaque base, on peut distinguer une face + et une face -. La face + est définie comme celle sur laquelle en parcourant le cycle à six atomes dans le sens des n croissants, on va également dans le sens trigonométrique (Fig. 4-1). On peut affecter à chaque base un vecteur perpendiculaire à son plan, et orienté de la face - vers la face +. Dans une paire de Watson-Crick, vue sur un plan, une base est + et l'autre est - ; mais il n'y a aucune raison stéréochimique générale pour laquelle des nucléotides dans l'eau interagissent de manière telle que le + soit associé au -.

Dans les cristaux, on trouve des appariements A.U de type ++ (Hoogsteen)⁽⁹⁸⁾ et de type +- (Hoogsteen renversé)⁽⁹⁹⁾. D'après la symétrie des fibres de poly A, on doit y trouver des associations A.A de type ++.

Orientation du sucre : l'orientation du sucre par rapport à la base à laquelle les cristallographes attachent beaucoup d'importance est plutôt une propriété métrique. Parler ici de configurations syn et anti ne ferait qu'obscurcir le tableau.

c) - Association de deux chaînes.

Considérons deux chaînes qui sont associées sur une partie de leur longueur par une interaction qui fait correspondre M_1 et M'_1 sur la première à M_2 et M'_2 sur la seconde. Soient $\vec{u}_1$ et $\vec{u}_2$ les vecteurs définissant l'orientation locale des chaînes aux points M_1 et M_2 . Formons les produits scalaires $\vec{u}_1, \vec{M_1M'_1}$ et $\vec{u}_2, \vec{M_2M'_2}$. Quand M'_1 tend vers M_1 et M'_2 vers M_2 , si ces produits sont de même signe, nous disons que l'association est parallèle. Sinon, disons qu'elle est anti-parallèle. Cette définition permet de compléter celle donnée dans la partie précédente qui ne pouvait pas s'appliquer aux associations de deux chaînes ouvertes. Nous allons maintenant montrer les liens qui existent entre nos définitions, la définition cristallographique et les idées intuitives relatives au parallélisme.

La définition ci-dessus peut être formulée de manière discrète dans le cas des chaînes ouvertes à sens de parcours unique. Donnons des numéros successifs aux points figurant des nucléotides successifs sur la chaîne en commençant par l'extrémité 5' libre. Si n (première chaîne) est associé à m (deuxième) et $n+1$ à $m+1$, l'association est parallèle. Si n correspondant à m , $n+1$ correspond à $m-1$, l'association est antiparallèle. Ou bien, on soude une extrémité 5' libre d'une chaîne à une extrémité 3' libre de l'autre, et on est ramené aux associations intrachaîne. Même schéma pour les associations de chaînes fermées : on coupe chaque chaîne de part et d'autre de l'association (en réalité une seule coupure suffit) et on joint une extrémité 5' libre à l'extrémité 3' libre de l'autre chaîne.

Considérons maintenant une structure hélicoïdale comme celle de l'ADN. Coupons-la par un plan perpendiculaire à l'axe de l'hélice. Soient M_1 et M_2 les intersections, $\vec{u}_1$ et $\vec{u}_2$ les vecteurs définissant l'orientation des chaînes, O le pied de l'axe de l'hélice et $\vec{n}$ le vecteur définissant la normale orientée au plan. On a $(\vec{n}_1, \vec{u}_1) (\vec{n}_2, \vec{u}_2) < 0$, mais cette relation n'est pas vraie pour n'importe quel plan passant par M_1 et M_2 . Elle requiert qu'on puisse privilégier un certain plan, donc elle n'est pas généralisable à n'importe quelle configuration. Elle a cependant quelque chose d'intéressant, c'est de répondre à une idée intuitive, celle d'une chaîne qui "monte" et d'une chaîne qui descend. On dit généralement que la chaîne qui monte est celle pour laquelle $\vec{n}_1, \vec{u}_1 > 0$; on peut aussi l'appeler chaîne +, l'autre s'appelant chaîne -.

Dans l'ADN, la chaîne + porte la base +, et la chaîne - la base -. Cette manière de décrire les choses peut s'étendre à des associations qui, sans être rigoureusement hélicoïdales, présentent une certaine régularité de ce type. On considère une relation qui au point M_1 de la chaîne 1 associe le point M_2 de la chaîne 2 et, en même temps, le point O d'une troisième chaîne, non orientée. On se limitera alors à considérer les sections par les plans OM_1M_2 . Le point O peut être, par exemple, le centre de gravité des bases, ou dans le cas d'un supercoil, le point mobile sur l'hélice principale.

Considérons toujours deux chaînes coupées par un plan, définissant par rapport à la normale au plan deux orientations locales. Si le résultat est +-, on obtient aussi +- en orientant la normale dans le sens opposé : il n'y a pas de manière privilégiée de regarder une association antiparallèle ; les deux filaments ne sont pas discernables, leurs signes s'échangent quand on change le sens de la normale au plan. Si par exemple, il s'agit de deux chaînes d'ADN, dont l'une est transcrite en ARN messenger, l'enzyme responsable de la transcription doit reconnaître une propriété autre que celles évoquées jusqu'ici, par exemple une séquence de bases.

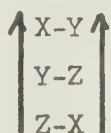
d) - Associations de trois chaînes.

Coupons trois filaments orientés par un plan dont la normale est orientée. Sur les trois signes déterminés à l'intersection, deux au moins sont pareils, + par exemple. Il existe alors deux possibilités pour l'association des trois chaînes : +++ ou ++-. Dans les deux cas, il existe une façon privilégiée de regarder les chaînes. Soient maintenant M_1, M_2, M_3 les intersections des chaînes avec le plan. Sur le cercle circonscrit au triangle $M_1M_2M_3$, la normale au plan induit un sens de parcours trigonométrique. Si nous sommes dans la situation ++-, on voit que chaque filament peut être individualisé. En effet, le - est d'abord reconnu du + comme celui qui n'est pas comme les deux autres, ensuite, les deux + peuvent être numérotés dans l'ordre où on les rencontre quand on parcourt le cercle dont il a été question, dans le sens trigonométrique, en partant de la chaîne -.

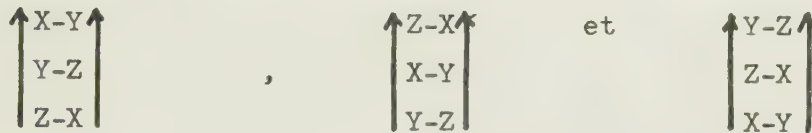
Dans la situation +++, il n'y a pas de filament privilégié, mais il y a deux types de relation d'un filament par rapport à l'autre. Supposons maintenant que sur les trois chaînes, une n'est pas orientable ; donnons-lui le signe 0. On a deux situations possibles : 0++ ou 0+- . Dans le premier cas, analogue à la situation ++-, les trois chaînes sont discernables. Dans le second, par contre, il est impossible de distinguer la chaîne + de la chaîne -. En effet, si on change le sens de la normale au plan, les signes + et - s'échangent, mais aussi le sens de parcours sur le cercle circonscrit aux trois intersections et par conséquent, les deux filaments s'échangent avec toutes leurs propriétés.

On voit, dans le cas de l'ADN que si on prend un troisième filament neutre, l'axe de l'hélice, on n'est pas plus avancé pour distinguer les chaînes. La disposition invariable des trois filaments est donnée sur la Fig. 4-m2. Les structures hélicoïdales à deux filaments parallèles peuvent être symétriques ou non. Dans le premier cas, par exemple celui du poly A, l'axe de l'hélice est axe de symétrie pour l'ensemble des deux filaments. Les seuls points invariants dans l'échange des filaments sont évidemment sur l'axe de l'hélice et ce sont aussi les seuls qu'on puisse discerner sans ambiguïté. Dans ce cas, 0, M_1 et M_2 sont alignés.

Une structure parallèle n'est pas forcément symétrique ; bien que jusqu'à présent les fibres de polynucléotides n'ont pas encore montré de structure parallèle non symétrique, le cas ne peut pas être exclu a priori. Une telle structure, dont les deux filaments seraient donc discernables, aurait des propriétés répliquatives différentes de celles de l'ADN. Si par exemple, les paires de bases non symétriques utilisées dans une telle structure sont X-Y, Y-Z et Z-X, l'association :



après séparation des filaments, pourra donner trois copies différentes :



de probabilités respectives $1/2$, $1/4$ et $1/4$, et dont seule la première est conforme à l'original.

Par conséquent, il ne faut pas sous-estimer le poids dû aux propriétés de symétrie des paires A.U et G.C dans l'explication des propriétés répliquatives de l'ADN. Cet aspect est parfois négligé par certains physicochimistes lesquels, inversement, surestiment la stabilité de la paire A.U.

e) - Relation entre le signe de la base et celui de la chaîne.

Dans les filaments d'ADN, les bases sont empilées de telle manière que la face + d'une base est en contact avec la face - de la base suivante. Cette loi est vraie pour la forme A comme pour la forme B ainsi que pour toutes les formes connues des fibres d'ARN. Si $\vec{v}_i$ est le vecteur unitaire définissant l'orientation de la base i , on peut définir la quantité :

$$\tau = \vec{v}_i \cdot \vec{v}_{i+1}$$

τ vaut 1 dans l'ADN de forme B. Il est peu différent de 1 dans l'ARN 10 et l'ADN de forme A. Une manière de voir si les bases sont à peu près orientées les unes par rapport aux autres, est de former les produits scalaires des paires de $\vec{v}_i, \vec{v}_{i+1}$ successifs. Pour une séquence de n bases, on peut construire la quantité :

$$\tau = \frac{1}{n-1} \sum_{i=1}^{n-1} \vec{v}_i \cdot \vec{v}_{i+1}$$

Pour un random-coil, structure que, à vrai dire, personne n'a vraiment caractérisé dans les acides nucléiques, $\tau = 0$. Une structure avec $\tau = 0$ a déjà été suggérée pour le poly C. En effet, le calcul des interactions entre bases dans la configuration de l'ADN indique une répulsion entre deux cytosines adjacentes. Les auteurs estiment qu'une configuration hélicoïdale dans laquelle une base sur deux est retournée serait plus stable⁽¹⁰¹⁾.

Une question physique qui mérite d'être posée est celle de la facilité avec laquelle une base peut se retourner dans une structure à un filament, et des éventuelles conséquences biologiques que ce degré de liberté pourrait entraîner.

Si on considère que le squelette sucre-phosphate est rigide, la question est celle des rotations de la base par rapport au sucre. Elle a fait l'objet de nombreux travaux de compilation. Les auteurs estiment généralement⁽¹⁰²⁾ que les purines peuvent facilement tourner autour de la liaison N9-C'1 tandis que pour les pyrimidines, il y aurait empêchement stérique. Cette vue, fondée sur des constructions et des simulations de modèles semble corroborée par le catalogue des structures cristallines de mononucléotides. Cependant, les expériences récentes tendent à prouver que les pyrimidines peuvent également tourner^(103, 104).

La part jouée par d'éventuels retournements de bases dans les processus de reconnaissance, et dans ceux de dénaturation-renaturation n'a pas été jusqu'à présent bien étudiée.

Supposons un instant que le retournement soit chose facile dans les structures d'un filament ; alors, les ARN messagers seraient des messages variables dans le temps, à huit ou six lettres suivant que la rotation est admise pour les pyrimidines ou non ; un code essentiellement basé sur une relation de complémentarité entre le codon et l'anticodon ne serait plus possible ; certains polynucléotides de Khorana devraient donner des résultats aberrants. Sinon, on devrait postuler l'existence d'un système qui, dans les ribosomes, rectifie les bases du codon.

Supposons maintenant que le retournement soit difficile à froid, mais qu'on y arrive en élevant la température. Alors, la renaturation devient un processus très délicat. Il serait intéressant de voir de quelle manière le poly U reconnaît à froid du poly A traité hermiquement.

Il existe des enzymes qui attaquent en principe des structures à un filament, tout d'abord suite après une base spécifique ; on devrait pouvoir, à l'aide de ces enzymes, étudier dans quelles conditions un polynucléotide est une séquence à six ou huit lettres plutôt qu'une séquence à quatre lettres. Lorsqu'un matériel présumé à un filament est mal dégradé par ces enzymes, les expérimentateurs en concluent souvent qu'il a des repliements secondaires.

Au moment de la synthèse, en tous cas, il est probable que l'enzyme de polymérisation rajoute les nucléotides dans la configuration usuelle exclusivement.

f) - Application au tARN.

L'étude des propriétés d'orientation des chaînes d'acide nucléique a été initialement entreprise pour répondre à la question suivante : y a-t-il moyen de concilier un arrangement en trèfle légèrement modifié, avec le modèle à trois filaments de Guschlbauer ? En effet, les expériences de diffusion centrale étaient compatibles avec un objet cylindrique

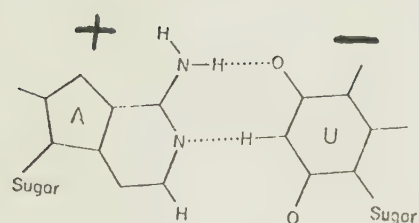


Fig. 4-1. Signes affectés aux bases. A gauche, une purine montre sa face + et à droite, une pyrimidine montre sa face - .

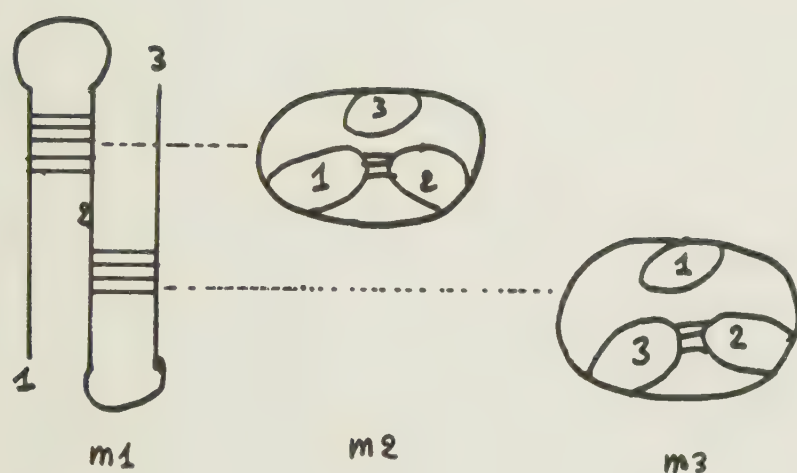


Fig. 4-m. Les schémas m_2 et m_3 expliquent l'irréalisabilité d'une structure régulière à trois filaments du genre montré en m_- , contenant des paires de Watson-Crick entre les filaments 1 et 2 d'une part, 2 et 3 d'autre part.

Tableau 4-I.

Bras Paires	GTTC	ACCEPTEUR	DIHU	ANTIC.	TOTAL
G.C	52	59	31	25	167
C.G	18	26	19	30	93
A.U	9	14	10	20	53
U.A	8	22	1	12	43

Evidemment, la totalisation des paires G.C n'a de sens que si on les définit de manière absolue. Le tableau ci-contre révèle la possibilité d'existence d'un phénomène hautement significatif. On peut se demander s'il existe une structure qui justifierait l'addition telle qu'elle a été réalisée. On voit en tous cas que l'étude des propriétés d'orientation peut conduire à des remarques intéressantes.

et allongé et, d'autre part, l'arrangement en trèfle commençait à se répandre.

Est-il possible de construire une structure à trois filaments régulière, dans laquelle des paires de Watson-Crick se forment quelque part entre les brins 1 et 2, et ailleurs entre les brins 2 et 3, comme cela est supposé sur la Figure 4-m₁ ?

Il est facile de démontrer que cela est impossible. Pour construire une telle structure à trois filaments, il faut qu'il se produise au centre, une inversion topologique de la permutation des filaments. D'une manière lapidaire, disons que pour ramener l'extrémité acceptrice près de l'anticodon au moyen d'une structure à trois filaments, il faut non pas deux boucles, mais trois.

Dans une structure à trois filaments qui ne sont pas tous parallèles, chaque filament peut être discerné. Des corrélations de séquence ont été recherchées pour voir si certaines "triades" apparaîtraient fréquemment. L'étude statistique a fait ressortir une autre propriété : la contrainte semblait plutôt porter sur la paire de Watson-Crick à l'intérieur de la triade.

Il est possible, pour chaque bras de la structure en trèfle, de faire une statistique des paires G.C, C.G, A.U et U.A. Il s'avère (tableau 4-1) que sauf pour le bras de l'anticodon, il y a une différence significative entre le nombre de paires G.C et celui des paires C.G. Naturellement, dans une structure à deux filaments régulière et symétrique, ces deux paires sont identiques. Par contre, G.C a un sens différent de C.G dans une structure à trois filaments, ou, de manière plus générale, quand il existe à côté de la structure à deux filaments, un élément d'orientation. On pourrait suggérer que dans le tARN, cet élément d'orientation est le fait que d'un côté de la région hélicoïdale et un seul, une boucle est présente : les deux extrémités sont physiquement discernables. Cependant, on ne verrait alors pas pourquoi le bras de l'anticodon échapperait à cet effet d'extrémité.

On peut, en second lieu, suggérer que les régions possédant ce phénomène de polarité sont en interaction avec d'autres parties de la même molécule ou bien avec d'autres molécules lors de la synthèse peptidique. Une interprétation satisfaisante du phénomène est alors difficile à fournir.

- CHAPITRE 5 -

PREPARATION DES ECHANTILLONS

Comme les expériences décrites se sont déroulées sur plusieurs années, avec des matériels divers, les techniques de préparation et de dosage utilisées sont variées. Nous nous sommes contentés d'appliquer des recettes connues, aussi n'est-il pas la peine de les exposer dans le détail.

a) ARN Ribosomal

Pour préparer l'ARN ribosomal, j'ai bénéficié de l'aide et des conseils de J.-Cl. Mounolou et de H. Fukuhara. La méthode d'extraction comporte une préculture en aérobiose de la souche YF237 suivie d'une culture en aérobiose. Le broyage est effectué dans un agitateur Nossal en présence de billes de verre. L'ARN est extrait par une suite de centrifugations fractionnées à petite, moyenne et grande vitesse, en présence de bentonite et de DNA-ase, puis par déprotéinisation au phénol. La solution obtenue est examinée en gradient de saccharose et montre l'existence de deux pics principaux à 16 et 23 S. La contamination par les ARN de faible poids moléculaire est faible. L'ARN préparé est conservé sous forme de précipité alcoolique à -20°C.

b) tARN total (Coli et levure)

Pour s'assurer que tout le tARN est dans le même état de charge, on a systématiquement déchargé le tARN total, soit de E.coli (Général Biochemicals) soit de levure (Boehringer) en suivant la méthode de Sarin et Zamecnik⁽¹⁹⁾. La décharge se fait à 37° dans une solution de 1,8 M de tris acétate à pH8. Les premières

expériences de diffusion centrale étaient faites avec le tARN total de E.coli déchargé, puis précipité à l'alcool. Pour les expériences qui ont suivi, le tARN de levure a été purifié par une procédure supplémentaire d'extraction au phénol-cétavlon qui nous a été communiquée par Kiselev.

Les dernières expériences utilisant du tARN total avaient pour objet l'étude de la diffusion aux grands angles par le tARN chargé ou non chargé.

Une des préparations, qui nous a été fournie par M. Yaniv, a été faite selon la méthode de Richardson⁽¹⁰⁵⁾. Le tARN de E.coli a été chargé avec un mélange de vingt acides aminés au moyen d'une fraction purifiée de protéines S-100. Le tARN chargé a été isolé par précipitations répétées à pH 5. Les deux formes (chargé et non chargé) ont été ensuite concentrées par ultrafiltration et dialysées longuement contre un tampon acétate de potassium à pH 7.

Une autre préparation, faite au laboratoire de J-P. Waller a été réalisée sur du tARN commercial de E.coli produit par Schwartz Bioresearch Inc. La charge a été réalisée avec le mélange des 20 acides aminés, en présence d'un extrait brut de E.coli.

L'efficacité de l'extrait a été vérifiée par la mesure de l'incorporation de valine marquée au C¹⁴. Ensuite, le tARN a été extrait du mélange par le phénol, précipité à l'alcool. La solution finale a été préparée par dialyse dans un appareil diaflo.

c) t-ARN^{Val} de E.coli

Le tARN^{Val} de E.coli a été préparé par M. Yaniv.

Le tARN total est d'abord enrichi en tARN^{Val} par un fractionnement comportant un passage sur colonne de DEAE cellulose benzoylée.

Le tARN est ensuite acylé avec sa synthétase pure, et le groupe amine libre est phénoxyacétylé puis isolé après une deuxième chromatographie sur DEAE cellulose. Une partie de ce tARN est déchargée dans un tampon tris pH9, précipitée à l'alcool, puis dissoute dans l'eau et dialysée extensivement contre de l'eau pure sous pression dans une cellule d'ultracentrifugation diaflo ; l'autre partie est directement dialysée contre de l'eau pure. Les contrôles ont montré que le tARN ne se décharge pas dans ces conditions. La pureté du tARN mesurée par son activité acceptrice était d'au moins 95 %. Pour obtenir des concentrations plus élevées (jusqu'à 5 %), la solution de tARN a été mise dans un sac à dialyse qui a été immergé dans de la poudre de Séphadex G-200.

d) Mesure des concentrations

Diverses méthodes ont été successivement employées, d'abord chimiques, puis optiques.

Phosphore. : Le microdosage du phosphore a été réalisé selon une méthode de minéralisation par voie humide et colorimétrie mise au point par les services de microanalyse du C.N.R.S. La minéralisation comporte une attaque sulfonitroperhydrique et conduit à la formation quantitative d'acide orthophosphorique. Cet acide est ensuite engagé dans un complexe phosphovanadomolybdique qui absorbe à 410 nm. La méthode convient pour des quantités de phosphore allant de 0,2 à 1 μ g.

Ribose. : L'ARN est précipité au TCA 5% à froid, le culot est hydrolysé au bain marie dans le même acide. A ce stade, on peut déjà déterminer la concentration en acide nucléique par la densité optique de l'hydrolysat. On prélève ensuite des quantités aliquotes sur lesquelles on fait agir une solution d' HCl concentré contenant de l'orcinol et du FeCl_3 . Il se forme un complexe qui absorbe à 650 nm. On peut ainsi doser 10 à 40 μ g de ribose.

Absorption U-V : Le coefficient d'absorption U-V à pH 7 dans NaCl 0,1 M. a été déterminé avec précision par A. Favre sur le tARN^{Val}. La valeur obtenue à 24°C. est de 7,4/micromole de nucléotide/cm. La précision est d'environ 3%.

e) Test de charge

En ce qui concerne le tARN^{Val}, l'état de charge ou l'activité acceptrice a été mesurée par M. Yaniv (voir plus haut). Sur du tARN total, les tests de charge ne peuvent être que très qualitatifs : on ne peut en pratique effectuer une charge quantitative que sur du tARN purifié, et avec une enzyme d'activation purifiée. J'ai essayé de voir si le tARN de levure utilisé acceptait la leucine marquée au C¹⁴. La préparation pH5 d'enzymes d'activation a été fournie par H. Fukuhara. Après quelques essais infructueux, des résultats positifs ont été obtenus en ajoutant du mercaptoéthanol dans le mélange réactionnel.

D'autres tests de charge, sur des échantillons commerciaux ont été effectués par la suite, soit dans le laboratoire de Mme Grunberg-Manago, soit dans celui de J-P Waller. Ces tests comprenaient une étape de filtration sur millipore.

f) Centrifugation analytique

Divers échantillons, tant de tARN total que de tARN^{Val} ont été soumis à un contrôle d'ultracentrifugation analytique, soit en optique Schlieren, soit en Archibald. J'ai bénéficié de l'aide et des conseils de J-Cl. Mounolou, Cl.Reiss et Mme M. Mossé. Un seul pic était observé, correspondant à une constante de sédimentation $S_{20,w}$ voisine de quatre.

CHAPITRE 6.

RESULTATS DES EXPERIENCES DE DIFFUSION CENTRALE DES RAYONS X SUR LES ACIDES RIBONUCLEIQUES.

Les premières expériences décrites ont été effectuées sur de l'ARN ribosomal ou du tARN total de levure ou de coli, en collimation linéaire infinie.

I. ARN RIBOSOMAL-EAU.

Pour me familiariser avec les techniques expérimentales, j'ai examiné le système ARN ribosomal-eau, déjà étudié par Timasheff, Witz et Luzzati⁽¹⁰⁷⁾ ; Luzzati, Witz et Timasheff⁽¹⁰⁸⁾ ; Witz⁽⁷⁾ . La courbe obtenue pour trois concentrations en solution aqueuse contenant 0,1 M de NaCl est donnée sur la fig. 6-a . Les résultats obtenus, toutes normations faites, sont très voisins de ceux trouvés précédemment. Ceux-ci avaient été interprétés à l'aide du modèle du "bâtonnet brisé" . Ce modèle, qui décrit des structures caténaïres, est constitué par une succession de cylindres de section, de masse et de longueur déterminées joints par des articulations flexibles. La comparaison des courbes expérimentales avec les courbes théoriques de divers "bâtonnets brisés" de section gaussienne permet d'extraire en principe trois paramètres : le rayon de gyration transversal, la masse par unité de longueur, et la longueur des bâtonnets élémentaires . Cependant, la précision des expériences ne permet pas d'avoir une détermination indépendante de chacun des paramètres. Pour un Rgt de 8,0 Å et des bâtonnets élémentaires longs de 100 Å , un rapport $A/c_e = 16,0$ avait été trouvé (il s'agit d'un rapport qui est relié à la masse par unité de longueur du bâtonnet ; son interprétation est donnée plus loin) . Dans les mêmes conditions, j'ai trouvé un rapport de 16,4 et toutefois, un accord un peu moins bon avec l'allure en bâtonnet brisé des courbes expérimentales.

Avec les valeurs communément admises à l'époque des expériences de Witz pour le volume partiel spécifique de l'ARN ribosomal ($0,56 \text{ cm}^3/\text{g}$) et de la distance, le long de l'axe d'une double-hélice d'ADN, entre deux plateaux de bases (3,4 Å) , ces expériences conduisaient à l'idée que l'ARN ribosomal devait dans sa quasi-totalité être en double-hélice, ce qui, vu sa composition chimique, exigeait que les paires

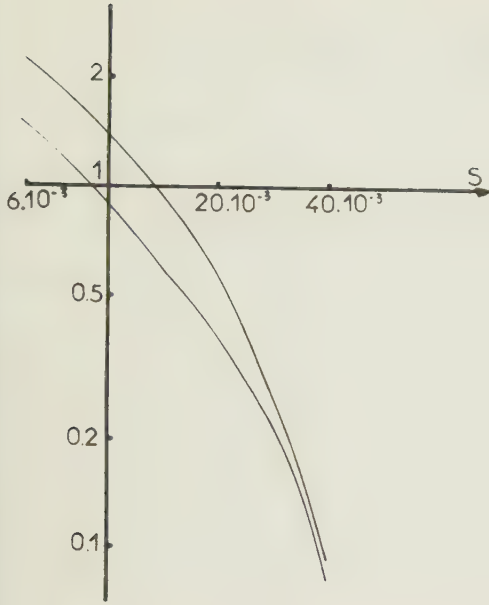


Fig. 6-a. Collimation linéaire infinie. La fonction portée en ordonnée est $j_n(s)/100c(I-\rho_0\psi)^2$. Courbe du haut: ARN ribosomal dans l'eau à 4°C. Courbe du bas: ARN ribosomal dans le glycol à la même température.

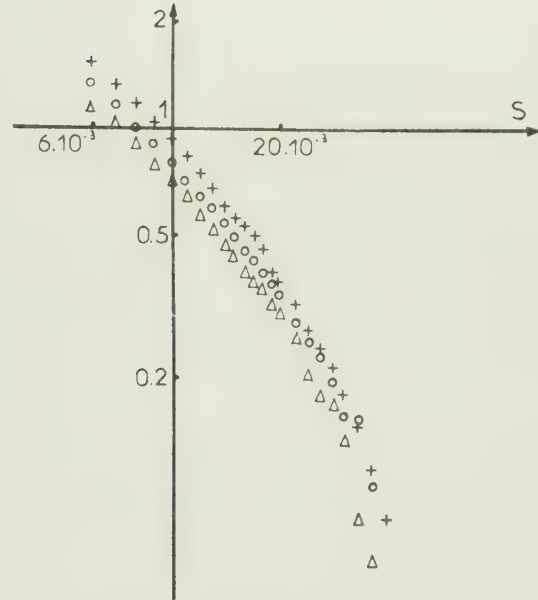


Fig. 6-b. L'ARN ribosomal dans le glycol à trois températures.

+ : 4°C.
o : 24°C.
Δ : 64°C.

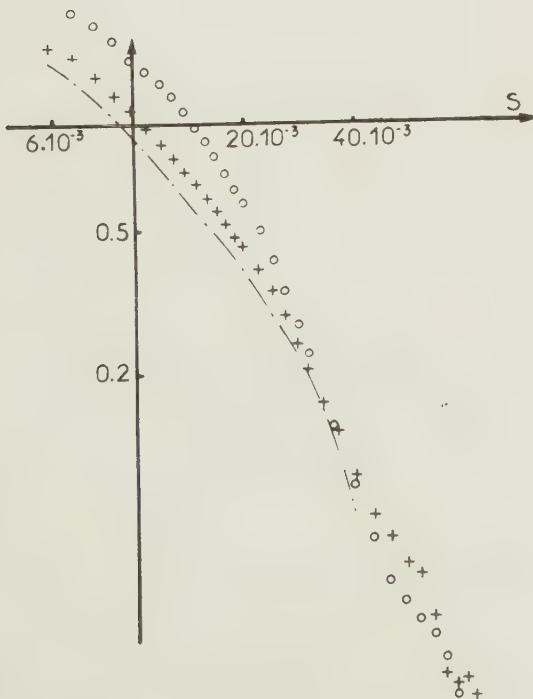


Fig. 6-c.

l'ARN de transfert dans l'eau à 4°C. (o) comparé au tARN dans le glycol à -21°C (+) et à l'ARN ribosomal dans le glycol à 4°C. A froid, l'effet relatif produit par le glycol sur l'ARN de transfert est très semblable à celui produit sur l'ARN ribosomal.

de bases ne soient pas exclusivement celles de Watson et Crick.

Les mêmes résultats, maintenant, seraient interprétés de manière différente. La valeur de $0,56 \text{ cm}^3/\text{g}$ pour le volume partiel spécifique pourrait être exagérée, et les déterminations les plus récentes effectuées sur de l'ARN de transfert varient entre $0,505^{(32)}$ et $0,530^{(109,110)} \text{ cm}^3/\text{g}$. Il se pourrait donc que la valeur utilisée pour l'ARN ribosomal soit trop élevée. D'autre part, la masse par unité de longueur d'un ARN en double-hélice est sans doute supérieure de plus de 10% à celle d'un ADN de forme B. Par conséquent, la fraction en double-hélice dans l'ARN ribosomal estimée dans les conditions précédentes était trop élevée. Si on tient compte de deux corrections possibles ($\bar{v}=0,53 \text{ cm}^3/\text{g}$; géométrie de l'ARN 10), la fraction en double-hélice estimée à partir des expériences passe de 99,5 % à 51 %.

De plus, le modèle du bâtonnet brisé était une première approximation; les expériences de dégradation enzymatique de l'ARN in situ dans les ribosomes^(III) suggèrent un arrangement différent des régions ordonnées avec, comme pour le tARN des segments hélicoïdaux, des boucles, et des sections de raccordement.

2. ARN RIBOSOMAL ET ARN DE TRANSFERT EN GLYCOL.

Les études de diffusion centrale des rayons X menées sur l'ADN dans l'éthylène-glycol^(II2-II4) avaient suggéré, contrairement aux idées courantes, que le glycol ne détruit pas facilement l'empilement des bases. La fusion de l'ADN dans ce solvant se fait à température plus basse que dans l'eau (vers 30°); elle fait passer la molécule d'un premier état ordonné à un deuxième état ordonné, correspondant vraisemblablement à des configurations en simple et double-hélice.

Nous avons tenté d'examiner si ces résultats pouvaient s'étendre aux acides ribonucléiques. Tant l'ARN ribosomal que le tARN ont été étudiés à différentes températures, dans le glycol. Les renseignements obtenus sont assez imprécis, pour les raisons suivantes :

1) Il semble que le magnésium, à des concentrations très faibles, en présence de glycol, intervient de manière curieuse sur la structure du tARN et encore plus sur celle de l'ARN ribosomal (voir à la fin de ce $\S$, les expériences optiques relatives à cet effet.)

2) Le contraste de densité électronique entre soluté et solvant est plus petit que dans le cas de l'eau.

3) La diffusion due au glycol devient importante dans la région des grands angles ($2\theta > 6^\circ$). D'autre part, il devient difficile de maîtriser, en collimation infinie, l'homogénéité du faisceau sur toute sa longueur, et certaines fluctuations du faisceau admissibles dans le cas d'un solvant "plat" comme l'eau ont des

répercussions plus graves dans le cas du glycol.

Comparaison de l'ARN ribosomal dans l'eau et le glycol à 4°C.

Si on compare les courbes obtenues à 4°C dans l'eau, et le glycol, on constate que dès cette température, le glycol a fortement influé sur la structure de l'ARN ribosomal, ce qui est bon accord avec l'idée que les régions en double-hélice dans l'ARN ribosomal sont relativement courtes. En s'aidant des abaques du bâtonnet brisé, on peut décrire chacune des courbes à l'aide de trois paramètres qu'on peut ensuite comparer. Dans le tableau ci-dessous, on exprime les résultats sous forme d'un ensemble de trois paramètres, qui ont une valeur comparative. Mais il faut bien se garder, en raison des nombreuses limitations, des expériences et de la méthode d'interprétation, de donner à ces paramètres une signification absolue. On a supposé que dans l'eau, comme pour le glycol, le volume partiel spécifique est de $0,53 \text{ cm}^3/\text{g}$. L'intensité diffusée mesurée est donnée sur la fig. 6-a .

	EAU	GLYCOL
Rayon de gyration transversal(A)	8,0	6,7
Masse par unité de longueur(el./A)	91	47
Longueur du bâtonnet élémentaire (A)	95	33

ARN ribosomal dans le glycol à différentes températures.

Les intensités absolues varient avec la température . Entre 4 et 64°C, on observe, toutes normations faites, une baisse de 35 % de l'intensité diffusée. Les courbes obtenues à 4, 24 et 64°C sont données sur la fig. 6-b. Les valeurs des paramètres qui ont été dégagées pour l'ARN ribosomal à 4°C sont compatibles avec une structure à un filament conservant l'empilement des bases. Si une légère séparation des bases s'opère à température plus élevée, le volume partiel spécifique pourrait augmenter. Une variation dans ce sens du volume partiel spécifique aurait pour corollaire qu'une partie de l'écart entre les courbes à 4° et à 64°C s'explique par la variation du volume partiel spécifique, et que par conséquent , la différence de 35% dans les intensités diffusées indique plutôt une borne supérieure de la différence des masses par unité de longueur.

Dans le domaine de températures considéré, nous n'avons pas trouvé d'argument

sérieux en faveur d'un changement conformationnel du même ordre que celui subi par l'ADN vers 30°C . Il n'est pas exclu qu'une transition importante a lieu à température plus basse.

t-ARN-GLYCOL A -21°C.

Une raison supplémentaire d'étudier le tARN dans le glycol à très basse température est que dans ces conditions, les nucléases sont inactives. On pouvait par conséquent espérer piéger un échantillon stable de tARN natif, sans avoir besoin de procéder à des purifications intensives. En fait, les intensités mesurées diffèrent de celles déterminées ultérieurement pour le tARN dans l'eau. La comparaison du tARN dans l'eau et le glycol avec l'ARN ribosomal dans l'eau ou le glycol montre que les intensités diffusées par chacune des espèces macromoléculaires dans l'eau et le glycol, sont dans le même rapport (fig. 6-c) .

En valeur absolue, les intensités déterminées dans l'eau pour les deux matériels sont voisines, de $s=0,008$ à $s=0,03 \text{ A}^{-1}$, et il en est de même des intensités diffusées dans le glycol. Par ailleurs, le tARN dans le glycol à -21°C présente une queue de courbe voisine de celle du tARN natif, ce qui impliquerait une certaine conservation de la structure hélicoïdale et empilée à courte échelle.

Expériences spectroscopiques.

En raison de la mauvaise reproductibilité des résultats obtenus, à température élevée, pour l'ARN de transfert dans le glycol, des expériences de dénaturation thermique ont été entreprises, en comparant les propriétés de l'ARN ribosomal et de l'ARN de transfert. Or, dans les deux cas, il était impossible d'obtenir une courbe de dénaturation thermique dans le glycol : il se produisait à chaque température des modifications lentes du spectre d'absorption U-V. Une étude un peu plus poussée a montré que les ions magnésium, à faible concentration, pouvaient être responsables du phénomène. En présence de 0,01 M de sulfate de magnésium, le spectre U-V des acides ribonucléiques évolue. La densité optique à 260 nm diminue, et au bout d'un temps assez long, il peut y avoir précipitation d'une partie de l'ARN . A température ordinaire, l'effet est plus marqué pour le tARN que pour l'ARN ribosomal. L'ADN dans les mêmes conditions, ne donne pas lieu au phénomène. La vitesse de la réaction croît avec la température, tant pour le tARN que pour l'ARN ribosomal , mais son étendue varie en sens inverse. Il s'agit donc d'une réaction endothermique, pratiquement irréversible, d'ailleurs, et qui rend difficile l'établissement de toute courbe de dénaturation thermique.

recouvre celle des très petits angles pour la diffusion des rayons X . La comparaison des courbes obtenues par les deux techniques laisse penser que la diffusion centrale des rayons X donne, pour ce polyélectrolyte, une intensité trop élevée à l'origine. Pour interpréter correctement le phénomène, il faudrait sans doute entrer dans le détail de la distribution des contre-ions autour du macroion que constitue l'ADN, tenir compte de la fraction des contre-ions osmotiquement libres ou liés ^(II7) . De plus, il faut préciser la relation entre le volume partiel spécifique utilisé dans certaines méthodes hydrodynamiques , et celui qui correspond aux équations de la diffusion centrale ^(II8) . Nous n'entrerons pas dans ces détails. Nous signalons simplement le problème pour justifier le fait que nous n'avons pas en réalité de point de comparaison parfaitement solide en ce qui concerne la diffusion des rayons X et qu'il faut provisoirement renoncer à étalonner le tARN par l'ADN.

Indiquons enfin qu'en présence de Mg 0,1 M, l'ADN paraît un peu plus compact, son Rgt passant à 8,0 Å . Cette diminution est d'autant plus significative que si on imagine des ions Mg à la périphérie de la molécule, en interaction avec les phosphates, le Rgt devrait au contraire augmenter. La forme en bâtonnet n'est pas altérée, et l'accord avec la courbe théorique du bâtonnet infini demeure excellent aux grandes valeurs de s. Dans une certaine mesure, d'ailleurs, cet accord est un coup de chance. En effet, l'intensité théorique diffusée par l'ADN de forme B présente un maximum vers $s=0,09$; dans la normalisation de l'intensité mesurée, une étape consiste, en collimation linéaire infinie, à soustraire la diffusion à $s=0,09$ de telle sorte qu'on obtient un signal "très pur", débarrassé des pics correspondant aux fluctuations à courte échelle de la densité électronique. Le cas des ARN serait différent.

t-ARN NON CHARGE

I) Choix de la concentration saline.

Witz avait montré dans son travail ⁽⁷⁾ que, à forte concentration saline (1M NaCl) , les risques d'agrégation du tARN sont notablement diminués. Il obtenait dans ce cas un Rg de 25 Å, et un poids moléculaire (déterminé par diffusion de la lumière) de 28800 . Nous avons également cherché à nous placer dans les conditions optimales de force ionique , qui se sont avérées différentes. Il semble que tant à 0,5M NaCl qu'à 0,3 M, une quantité importante d'agrégats est présente. A ces molarités de chlorure de sodium, il a été impossible d'extraire un rayon de gyration. Un effet d'agrégation du tARN aux concentrations élevées de NaCl a également été observé par Millar ^(II8b) par diffusion de la lumière

Par ailleurs, en opérant à force ionique trop faible (0,01 M), les effets de corrélation entre les orientations des molécules en solution risquaient de devenir importants, et gêner l'interprétation des résultats.

La concentration de 0,1 M NaCl s'est avérée convenable, et toutes les études ultérieures ont été entreprises à cette molarité.

2) tARN non chargé, conditions standard.

La courbe extrapolée à concentration nulle, pour les solutions de tARN de levure dans NaCl 0,1 M et tris pH 7 0,01M est donnée sur la fig. 6-d. Les extrapolations relatives au poids moléculaire apparent ainsi qu'au R_g apparent sont données sur la fig. 6-f. On voit apparaître dans la région des $s=0,06$ un plateau tout à fait caractéristique, et reproductible d'une série d'expériences à l'autre.

Jusque là, les expériences faites en collimation linéaire infinie avaient été poussées jusqu'à $s=0,03$ ou $s=0,04$. Dans nos expériences, on commençait à obtenir un signal en dehors de la région précédemment explorée. Pour interpréter la courbe obtenue, le modèle du bâtonnet brisé devenait insuffisant, et cela nous a conduit à mettre au point les méthodes de simulation des modèles (Chapitre 3).

De la courbe obtenue, on pouvait extraire plusieurs paramètres :

-Rayon de gyration apparent	: 22,3 Å
-Rayon de gyration vrai	: 23,5 Å
-Volume	: 41000 Å ³
- $j_n(0)/c_e (I-\rho\psi)^2$	: 288,8
-Poids moléculaire ($\bar{v}=0,53$)	: 25900

D'autre part, si on assimile la molécule à un ellipsoïde de révolution allongé, dont le volume et le R_g sont donnés par l'expérience, on trouve un rapport grand axe/petit axe de 3,6 et si on assimile la molécule à un cylindre de révolution défini par les mêmes paramètres expérimentaux, on trouve un rayon de 12,9 Å pour une hauteur de 78 Å.

3) Etude du rapport surface/volume.

Un trait qui distinguait le tARN des autres acides nucléiques étudiés jusqu'alors était la présence, aux angles moyens (c'est-à-dire, ici, dans la région des s compris entre 0,02 et 0,06), d'une allure asymptotique en I/s^3 (correspondant, en collimation ponctuelle à une allure en I/s^4). Une telle allure suggère l'emploi des modèles à deux densités de la distribution électronique

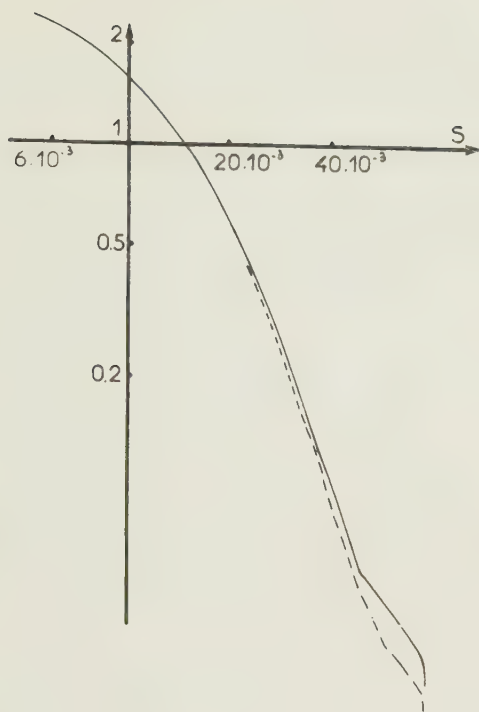


Fig. 6-d. Collimation linéaire infinie (suite). En traits pleins la courbe obtenue pour le tARN total non chargé de levure, comme d'E. coli. En traitillé, la courbe obtenue par Lake et Beeman, transformée en collimation infinie.

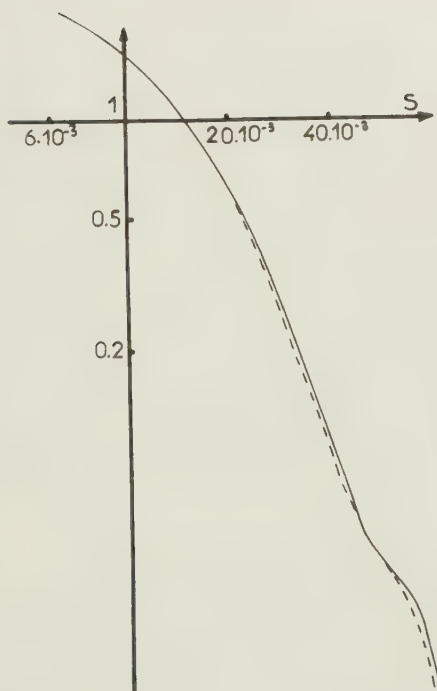


Fig. 6-e. Effet du magnésium externe sur la structure. Le magnésium, en faibles quantités, semble avoir un léger effet sur la structure secondaire. En traitillé, la courbe pour le tARN total de levure, en présence de magnésium, est comparée à celle pour le tARN total privé de Mg externe par un traitement à l'E.D.T.A.

en solution (voir Chapitre 2) , et permettant de calculer un rapport surface/volume. Nous avons cherché à savoir s'il était possible de tirer parti de cette valeur pour obtenir quelques renseignements complémentaires concernant la géométrie des régions hélicoïdales dans la molécule. Nous avons ainsi étudié, en ordinateur, l'allure asymptotique des courbes diffusées par différents modèles de polynucléotides. Il n'a pas semblé que l'exploitation de ce paramètre coulait de source.

A la même époque, aux Etats-Unis, Lake venait d'achever son étude sur le tARN total de levure, par diffusion centrale des rayons X⁽¹³⁾ , avec une collimation de type rectangulaire, qui permettait d'obtenir avec précision les intensités diffusées jusqu'à $2\theta = 10^\circ$. Des études de construction de modèles suivies de simulation en ordinateur l'avaient conduit à rejeter les arrangements ouverts du trèfle, et leur préférer les arrangements compacts 4-4 ou 6-2 .

Il nous est apparu tout de suite que les résultats de Lake étaient en bon accord avec les nôtres, mais que nous ne pouvions partager son interprétation (voir Chapitre 3) .

La possibilité d'obtenir des détails sur la structure fine, la précision supérieure obtenue par Lake, puis la proposition faite par Yaniv de nous fournir un accepteur purifié, ont été des raisons décisives pour abandonner la collimation linéaire infinie au profit de la collimation linéaire finie.

4) Effet de la température.

Une expérience portant sur trois concentrations de tARN de E.coli montre qu'en passant de 4°C à 40°C, le rayon de gyration subit une petite augmentation, sans doute inférieure à 6% . On peut donc conclure à une légère dilatation entre ces deux températures physiologiques extrêmes : plus précisément, les plateaux de bases ne sont jamais parfaitement empilés, ils s'écartent et se rapprochent ; il n'est pas exclu que la différence observée entre les R_g à 4 et 40°C reflète simplement une différence entre les états moyens de structures dynamiques dont la plupart des états sont communs.

5) Effet du magnésium.

En 1966, les expériences de Fresco⁽¹⁴⁾ sur la dénaturation thermique de l'accepteur purifié tARN^{Leu} , en présence ou en absence de magnésium, avaient frappé les imaginations. Si on incube le tARN^{Leu} au-delà de 60°C en présence de versène, et qu'on le refroidit, on le piège dans une forme biologiquement inactive, même en présence de Mg ajouté à froid. Pour corrélérer d'éventuels changements d'état physiologique, avec des modifications de la structure tertiaire, les quatre solutions suivantes ont été préparées :

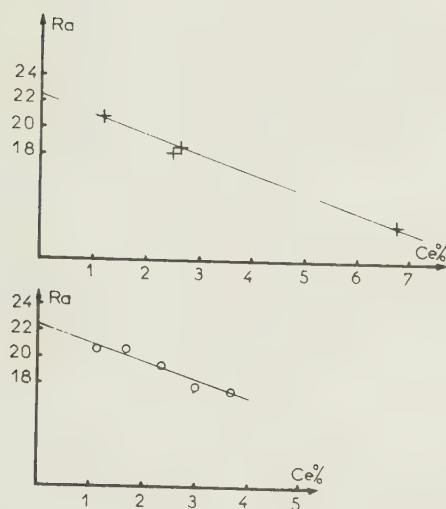


Fig. 6-f. Extrapolation à concentration nulle du R_g apparent pour le tARN total de levure (+) ou de E. coli (o)

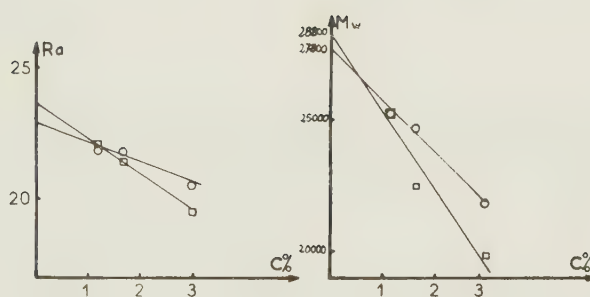


Fig. 6-g. Effet de la température sur le rayon de gyration et le poids moléculaire apparents du tARN total de E. coli. Ronds : mesures à 4°C. Carrés : mesures à 40°C.

a) tARN non chargé, privé de magnésium "externe" par un traitement au versène à 20°C, dans NaCl 0,1 M et tris pH 7, 0,01 M.

b) même préparation, mais avec 0,01 M $MgCl_2$ rajouté ultérieurement.

c) tARN non chargé, privé de magnésium "interne" par un traitement au versène à 60°C, dans le même solvant que a) .

d) même préparation que c) , mais avec 0,01 M $MgCl_2$ rajouté ultérieurement.

La conclusion de cette étude correspond assez bien à celle que nous avons tirée du système ADN-Mg : l'addition de faibles quantités de magnésium externe ne change pas de manière appréciable la géométrie du tARN . Les courbes obtenues en b) sont presque identiques aux courbes obtenues en a). De même, les courbes obtenues dans les cas c) et d) sont très voisines. Par contre, la privation de magnésium interne affecte la forme de l'intensité diffusée : la bosse aux grands angles à tendance à s'escamoter en l'absence de Mg interne, tandis qu'on ne voit toujours pas apparaître, vers les angles plus petits, l'épaule caractéristique des modèles 4-4 ou 6-2 . On est alors tenté de conclure qu'en absence de Mg interne, la molécule conserve une forme compacte et allongée.

Les mesures les plus complètes concernent la comparaison de a) et b). Le rayon de gyration est très sensiblement le même, l'intensité diffusée, après normation est très semblable dans les deux cas (Fig. 6-e) , mais on peut noter un léger écart aux grands angles, écart notablement amplifié si on considère l'intensité diffusée avant l'étape de normation qui consiste à soustraire la contribution aux grands angles ($s=0,09$). Nous reviendrons sur ce point plus loin.

Toutes les expériences qui suivent ont été réalisées en collimation linéaire finie. Les résultats sont exploités après corrections de collimations.

PHENOXY-ACETYL tARN^{Val}_I

L'ARN de transfert chargé est instable à pH 7, et le détachement de l'acide aminé peut se produire en un temps qui varie de quelques minutes à quelques jours selon les cas. Le meilleur moyen d'étudier un tARN dont on soit sûr qu'il est chargé à 100% est d'utiliser une espèce dans laquelle l'acide aminé a été phénoxyacétylé sur son groupe amine. Un tel tARN est séparable de la forme non chargée par chromatographie (Ch. 5). C'est pourquoi il a paru intéressant de travailler avec ce matériel. Les résultats précédents sur le tARN non chargé, en collimation infinie semblaient indiquer que la molécule n'adopte aucune des configurations tenues jusqu'alors pour vraisemblables . On s'est demandé si la forme chargée ne présenterait pas, elle, l'épaule à $s=0,03$ caractéristique

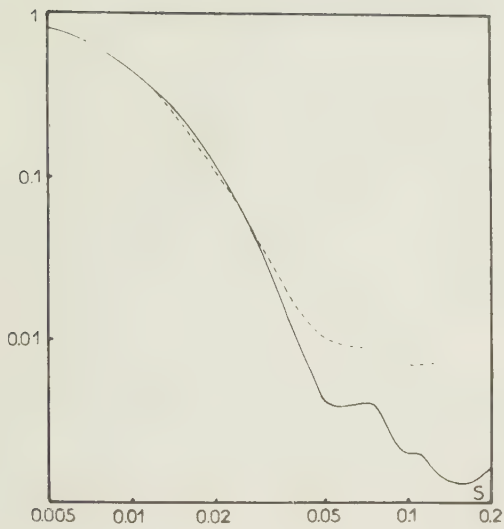


Fig. 6-h. Collimation linéaire finie. Intensité diffusée par le phénoxyacetyl tARN^{Val} (traitillé) ou le tARN^{Val} non chargé (trait continu), ramenée à la collimation ponctuelle, et normée en sorte que $I(0) = I$.

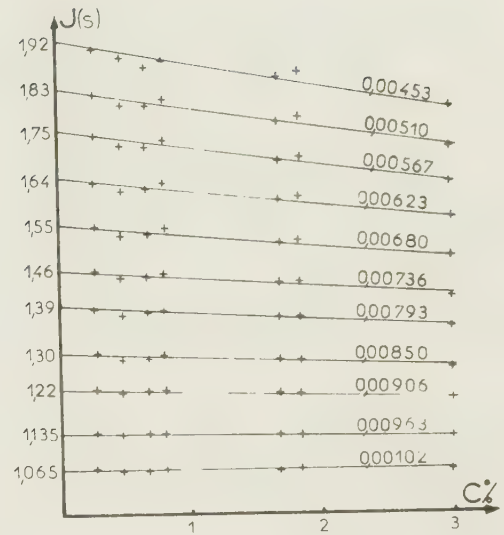


Fig. 6-i. Phénoxyacetyl-tARN^{Val}. Extrapolation des intensités mesurées vers la concentration nulle, aux valeurs de s indiquées sur les droites, avant correction de collimation. Echelle arbitraire.

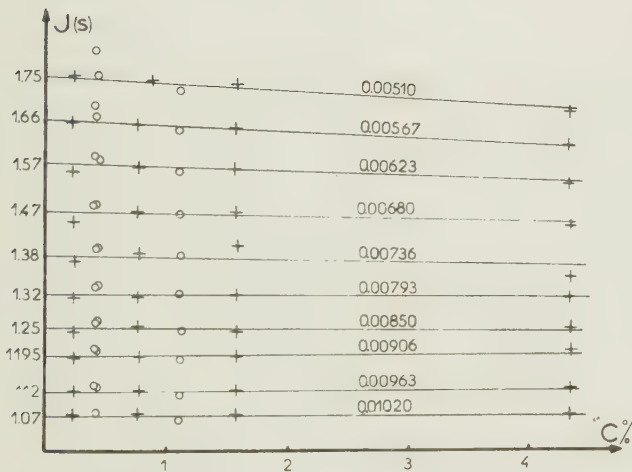


Fig. 6-j. tARN^{Val} non chargé, irradié (o) ou non irradié (+). Extrapolation des intensités à concentration nulle comme pour la fig. 6-i.

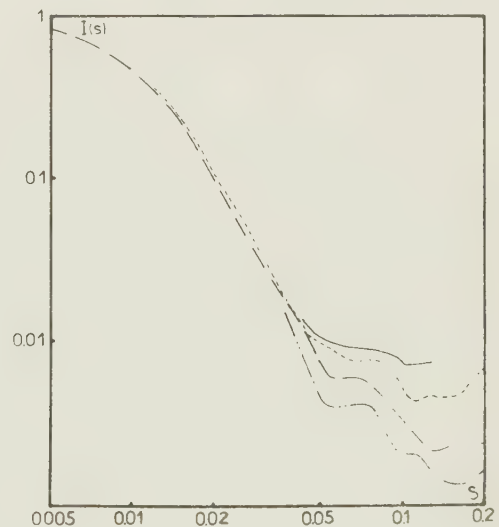


Fig. 6-k. Problème des niveaux. Un modèle de tARN calculé par la méthode des cubes (----) et par celle de Langridge (---), comparé aux grands angles avec les courbes expérimentales.

de toutes les variations classiques du trèfle. Des échantillons de tARN phénoxy-acétylés ont été étudiés. Le groupement phénoxy est fixé sur l'amine libre de la valine, on a donc un composé du type peptidyl -tARN .

Pour l'extrapolation à concentration nulle, sept concentrations ont été utilisées (fig. 6- i) allant de 0,3 à 3% . Aux grands angles, deux concentrations ont été utilisées : 2 et 4 % .Le contraste dans la région des $s=0,05$ ou même des $s=0,1 \text{ \AA}^{-1}$ est très bon. Au-delà de $s=0,15$ par contre, la diffusion de l'eau devient prépondérante.

La courbe complète, après corrections de collimation, pour le phénoxyacetyl-tARN^{Val}_I est donnée sur la fig. 6- h . L'interprétation des résultats est faite plus loin.

$$\frac{\text{tARN}^{\text{Val}}_{\text{I}} \text{ NON CHARGE} - \text{tARN}^{\text{Val}}_{\text{I}} \text{ IRRADIE.}}{\text{I}}$$

L'étude du phénoxyacetyl-tARN^{Val} n'avait pas permis de mettre en évidence une structure de type 4-4 ou 6-2 . On pouvait penser alors que les différences entre tARN chargé et non chargé seraient faibles. D'autre part, Yaniv Favre et Barrell venaient d'obtenir leur résultat sur la liaison thioU-C . J'ai donc étudié, toujours à 4°C et dans le même solvant, le tARN^{Val}, irradié et non irradié.

L'irradiation à 330 nm semble produire peu d'effet sur la structure du tARN telle qu'elle apparaît aux rayons X. On voit, sur l'extrapolation à concentration nulle des points près de l'origine (fig. 6- j) que les trois concentrations utilisées pour le tARN irradié se placent assez bien sur les droites correspondant aux quatre concentrations de tARN non irradié, avec cependant, une petite tendance possible à l'aggrégation. Les queues de courbe sont pratiquement identiques. Après corrections de collimation, la courbe de diffusion centrale des rayons X par le tARN^{Val}, non chargé, irradié ou non, est donnée sur la fig. 6- h . La différence principale avec la forme phénoxy réside dans les niveaux de la diffusion aux grands angles.

DETERMINATION DU POIDS MOLECULAIRE.

La structure chimique qu tARN^{Val} étant établie, la détermination du poids moléculaire a surtout valeur de contrôle.

L'intensité à l'origine est liée au nombre N d'électrons de la molécule par la formule :

$$i_n(0) = N c_e (1 - \rho_0 \psi)^2$$

Quand le volume partiel spécifique de la molécule est connu, on peut calculer N . Les déterminations récentes de volume partiel spécifique pour le tARN sont divergentes ^(32,109,110) et vont de 0,5 à 0,53 comme il a été dit. Ceci donne une première incertitude sur le poids moléculaire.

Le poids moléculaire est lié à N par une relation $M = k N$, où k est une constante voisine de 2 et qui dépend de la composition chimique du corps étudié. Pour le sel de sodium d'un acide ribonucléique, k est très voisin de 1,966. Cette valeur a été calculée sur la séquence connue du tARN^{Val}.

En pratique, il y a souvent une hypothèse implicite concernant le poids moléculaire quand on détermine les concentrations. Par exemple, dans le dépouillement des résultats expérimentaux, les concentrations ont été déduites des mesures d'absorption U-V. des échantillons, avec le coefficient $E_{1cm}^{2\%} = 232,4$. Si, en moyenne, comme l'a déterminé Favre, une mole par litre de nucléotide de tARN^{Val} a un coefficient d'absorption de 7,4 à la même longueur d'onde (260 nm), le poids moléculaire implicite est de 24200, correspondant à 12310 électrons.

Si, utilisant ce poids moléculaire implicite M_I dans une détermination des concentrations, on trouve par application de l'équation I un poids moléculaire M_2 , alors les expériences de rayons X sont compatibles avec un poids moléculaire $M = \sqrt{M_I M_2}$.

Dans le cas du tARN^{Val}, on trouve de cette manière, dans l'état non chargé un nombre d'électrons de 13790 si on choisit 0,53 pour volume partiel spécifique, et un nombre plus faible si $\bar{v} = 0,5$. Cette valeur est à comparer à celle de 13366 électrons qu'on obtient en considérant 76 nucléotides plus 76 atomes de sodium. Le poids moléculaire correspondant à ce nombre d'électrons est de 27100. Dans le cas de la forme phénoxy, si on admet en moyenne un coefficient d'absorption de 7,5 par mole par litre de nucléotide (ce qui revient à considérer que le groupement phénoxy absorbe autant qu'un nucléotide), on en déduit un nombre d'électrons de 13930 et un poids moléculaire de 27400. Par conséquent, s'il y a agrégation des molécules, elle ne peut être que très limitée.

DETERMINATION DU RAYON DE GYRATION.

Les corrections de collimation ont été appliquées sur la courbe expérimentale extrapolée à concentration nulle. Le programme utilisé place obligatoirement les points proches de l'origine sur une gaussienne. Les valeurs de R_g ainsi obtenues

sont respectivement de 24,7 Å pour la forme non chargée et de 24,9 Å pour la forme phénoxy . La différence observée est minime et non significative.

D'après la détermination du poids moléculaire, il est peu probable que des agrégats entachent d'erreur les valeurs de R_g . Des contrôles faits en centrifugation analytique n'indiquent pas la présence de plus d'une espèce moléculaire. Cependant, on peut évaluer dans quelle mesure il faudrait réviser les valeurs de R_g si des dimères étaient présents.

Considérons un mélange de monomères ayant un R_g de 23,5 Å avec des dimères ayant un R_g de 33,0 Å dans la proportion de 9 molécules de monomère pour une molécule de dimère. Si on calcule l'intensité diffusée par le mélange près de l'origine en ajoutant les contributions des deux espèces, et qu'on l'analyse suivant la loi de Guinier, on peut en extraire un R_g apparent de 24,8 Å et un poids moléculaire supérieur de 8% à celui du monomère.

Si on estime que les 3% de différence entre le poids moléculaire mesuré en utilisant $\bar{v}=0,53$ et celui calculé par les données chimiques est dû à la présence d'une petite quantité de dimères, il faudrait ramener la valeur du R_g du tARN^{Val} de 24,7 à 24,2 Å . Si on utilise $\bar{v}=0,50$ cm³/g pour le volume partiel spécifique, le poids moléculaire déterminé par diffusion centrale est plus petit que celui déterminé par voie chimique. Là encore, on se heurte à la définition exacte de la portion de solution qui constitue la particule vue par les rayons X .

PARAMETRES DE FORME.

En calculant les paramètres macroscopiques qui permettent de se faire une idée approximative de la forme de la molécule : volume, rapport surface/volume, R_g , paramètres de l'ellipsoïde ou du cylindre équivalent, on trouve des valeurs très voisines pour les deux formes étudiées du tARN^{Val} . Les diverses quantités évaluées sont portées dans le tableau ci-dessous.

	Phénoxyacetyl tARN ^{Val}	tARN ^{Val} non chargé
Rayon de gyration électronique(Å).	24,9	24,7
Poids moléculaire si $\bar{v}=0,53$	27400	27100
si $\bar{v}=0,50$	25800	25500
Volume (.10 ⁴ Å ³)	4,20	3,95
Surface (.10 ⁴ Å ²)	1,16	1,04
Ellipsoïde équivalent rapport grd.axe/petit	3,7	3,8

QUEUES DE COURBES.

le résultat le plus surprenant de cette étude est la différence importante dans la diffusion aux grands angles par les deux systèmes. Ces différences sont tout à fait significatives.

Les courbes obtenues pour le tARN non chargé sont en excellent accord avec celles obtenues par d'autres équipes^(13,44,119). Les intensités mesurées aux grands angles ont été les mêmes, qu'il s'agisse du tARN^{Val} non chargé, irradié ou non, ou du tARN total, déchargé, de levure.

Par ailleurs, le niveau a été mesuré avec précision dans le cas de la forme phénoxy ; la diffusion dans la région des $2\theta = 8^\circ$ est triple de celle du solvant. Cependant, les intensités théoriques pour différents modèles de tARN, calculées par la méthode de Langridge étaient nettement en-dessous des valeurs expérimentales. Cela est curieux car on comprend mal qu'une molécule réelle présente apparemment plus d'ordre que la molécule simulée en ordinateur avec ses tronçons parfaitement hélicoïdaux. Ceci nous a conduit à revoir les procédés de calcul des intensités théoriques (Ch. 2). La situation des courbes expérimentales par rapport aux courbes théoriques calculées soit par la méthode des cubes, soit par la méthode de Langridge mérite d'être comparée à celle qui s'est présentée dans le cas de l'ADN.

La méthode des cubes donne une intensité théorique un peu plus faible que l'intensité expérimentale pour l'ADN déterminée par Bram et Beeman. En ce qui concerne le tARN, elle est dans le même rapport, avec la courbe expérimentale pour la forme phénoxy. Par ailleurs, elle donne une intensité théorique un peu supérieure à l'intensité expérimentale pour l'ADN déterminée par Kirste et Oberthür, et très supérieure à l'intensité expérimentale pour le tARN non chargé (comparer les fig. 2-e et 6-k).

La méthode de Langridge donne une intensité théorique trop faible, dans le cas de l'ADN, qu'il s'agisse des courbes de Bram ou d'Oberthür. Par contre, elle donne une intensité supérieure à celle du tARN non chargé.

De cet ensemble de comparaisons, on peut dégager l'idée que c'est plutôt le tARN non chargé qui a un comportement anormal aux grands angles. La question se pose donc de savoir quelles sont les particularités structurales qui peuvent rendre compte du faible niveau de l'intensité diffusée par le tARN non chargé

aux grands angles.

Une représentation qui permet de jeter une lumière sur ce problème a été récemment proposée dans une étude du poly C ^(I20). La chaîne polynucléotidique est assimilée au produit de composition d'un fil de forme quelconque et d'une gaussienne. Le niveau de diffusion aux grands angles dépend essentiellement de la gaussienne. On pourrait alors imaginer que la différence des niveaux observée entre le tARN chargé et non chargé tient à l'inclinaison des bases par rapport à la chaîne des phosphates. Cependant, dans la limite des "géométries" classiques (ADN, ARN IO, etc.) , le niveau de l'intensité diffusée aux grands angles varie peu (voir fig. 3-b).

On peut aussi tenter d'expliquer les différences ~~en~~ termes de désordre. Passer d'une forme à l'autre exigerait des perturbations profondes de la structure secondaire, brouillant les atomes des 70 % de la molécule avec une amplitude de l'ordre de 5 Å . Cependant, il est extrêmement difficile de simuler en ordinateur un modèle qui tienne compte des liaisons hydrogène , de la structure tertiaire, et qui en même temps soit désordonné. Il y aurait d'ailleurs quelques difficultés à corrélérer ce désordre avec les données expérimentales apportées par d'autres techniques.

S'il existe plusieurs critères expérimentaux qui permettent de distinguer le tARN chargé du tARN non chargé ^(I5, I8, I9, 20, 95) , les expériences optiques n'ont fait apparaître jusqu'à présent que des différences peu significatives, que ce soit par dichroïsme circulaire ^(I22) ou par dénaturation thermique ^(I23).

Un point semble acquis, cependant, c'est que le niveau de l'intensité diffusée aux grands angles dépend de la concentration en magnésium. Ceci a été montré sur une série additionnelle d'expériences, portant sur du tARN total de coli, chargé et non chargé à pH 5 pour les deux formes et à pH 7 également pour la forme non chargée (dans NaCl 0,1 M) . Dans les deux systèmes, l'intensité diffusée aux grands angles baisse quand on rajoute du magnésium . Cependant, on ne peut pas aller jusqu'à 10^{-1} M de magnésium car le tARN commence à précipiter. On peut suggérer que lorsque le magnésium se fixe sur une région hélicoïdale, il la stabilise, mais en même temps, il la déforme.

Enfin, on ne peut pas exclure la possibilité que ces différences observées aux grands angles correspondent à des changements petits, mais considérablement amplifiés pour des raisons techniques encore mal élucidées . Il faudra attendre d'autres travaux, d'autres déterminations d'intensités aux grands angles, pour que la diffusion des rayons X aux petits angles connaisse aussi l'échelle absolue aux angles plus grands que ceux qu'elle a coutume d'étudier.

- CHAPITRE 7 -

- LE MODELE 4-2-2. -

Premier modèle

Disposant des données de rayons X et de la connaissance de l'interaction thioU-C, j'ai procédé avec A.Favre et M. Yaniv à une étude de modèles. Pour commencer, nous avons pris le segment ordonné de quatre paires de bases DiHU, connecté aux deux bases qui le prolongent vers l'extrémité 5' : l'adénine 9 et le thio uracile 8. En conservant au segment DiHU une structure secondaire voisine de celle de l'ARN 10, on peut envisager deux modes de repliement du segment des résidus 8-9 : soit dans le petit, soit dans le grand sillon du tronçon hélicoïdal. Il semble à peu près exclu que le repliement dans le petit sillon permette un voisinage suffisant du thiouracile et de la cytosine 13. Le segment 8-9 a donc été logé dans le grand sillon du bras DiHU. La position du thioU a été choisie de manière à faciliter la liaison chimique avec C 13. Selon Favre et Yaniv, la position optimum se trouve empilée au-dessus de la paire C.G plutôt que d'être dans le plan de cette paire.

Partant de là, il devenait possible de construire une molécule très asymétrique, en introduisant un minimum de torsions dans la chaîne des phosphates :

- 1) le bras de l'anticodon se place naturellement en dessous du bras DiHU.
- 2) le bras accepteur se place au-dessus du bras DiHU, dans le prolongement du segment 8-9.

.../...

Par contre, la position du bras GTWC ne s'imposait pas de manière évidente. La volonté d'obtenir un rayon de gyration compatible avec les données de rayons X conduisait à placer ce bras au même niveau que le bras accepteur, selon un modèle 4-2-2. Dans l'autre famille d'arrangements, celle des modèles 2-4-2, et toutes choses égales par ailleurs, le Rg est plus petit d'au moins deux Å. Un modèle 2-4-2 a d'ailleurs été construit avec Yaniv et Favre, sur la suggestion de Crick, puis un autre avec Danchin⁽⁴⁵⁾ (Ch. 3).

Nous avons construit plusieurs variétés du modèle 4-2-2.

Un premier, très compact, avec les coordonnées de l'ARN 11, et des interactions étroites entre les différentes parties de la molécule, a donné un rayon de gyration de 21,5 Å. En d'autres termes, si on prend l'arrangement le plus asymétrique, mais construit de manière compacte, on est bien en dessous du Rg expérimental.

Les modèles suivants ont été plus aérés ; en particulier, nous n'avons pas cherché à faire interagir les boucles, comme c'est le cas dans le modèle de Cramer⁽⁶⁴⁾. Nous avons cependant vérifié que certaines interactions supplémentaires entre différentes régions de la molécule demeuraient possibles, sans nous prononcer sur leur réalité. La source la plus grande de variabilité est dans les positions relatives des bras GTWC et accepteur. Dans les premiers modèles, cette imprécision se répercutait sur la position du doigt qui se trouvait plus ou moins flotter à l'écart au corps principal. Finalement, nous nous sommes arrêtés sur un modèle dans lequel le tronçon DiHU était entouré par le doigt, solution présentant plusieurs avantages :

- empilement convenable des nucléotides dans la région du doigt
- celui-ci fait corps avec le reste de la molécule
- positions relatives des bras GTWC et accepteur relativement bien définies.
- possibilité (intéressante d'un point de vue évolutif) de passer facilement à un modèle 3-2-3.

Second modèle 4-2-2

Après la publication du premier modèle 4-2-2, deux données expérimentales ont permis de préciser les hypothèses sur la structure du tARN.

La première est la constatation faite par Levitt⁽⁴³⁾ de l'existence d'une corrélation entre une base de la boucle DiHu et la première base du doigt. L'ensemble forme soit une paire G.C, soit une paire A.U. Dans le cas du tARN^{Val} de E.coli, G occupe la position 15, et C la position 48.

La deuxième est la convergence des données biochimiques relatives à la réactivité des bases dans la boucle GTTC. Il apparaît de plus en plus que la plupart des bases de cette boucle sont peu accessibles à des agents chimiques divers. Evidemment, on ne peut pas se contenter de formuler un jugement sur une donnée chimique isolée, puisqu'il existe certaines contradictions entre les données de réactivité des bases, et celles de dégradation enzymatique. Cependant, on peut dire que, dans l'ensemble, la boucle GTTC est la moins accessible des trois boucles du tARN, et cet aspect n'avait pas été incorporé dans le premier modèle.

Je me suis donc intéressé d'abord à la contrainte de séquence de Levitt. Celle-ci peut évidemment être incorporée dans un modèle 2-4-2, et j'ai vérifié de plus avec Danchin qu'elle était compatible avec un modèle semblable à celui de Connors⁽⁴⁴⁾. Cependant, la molécule présente un aspect tourmenté et on se demande si dans ce genre de situations, le gain d'énergie libre obtenu par les liaisons hydrogène n'est pas amplement perdu par la torsion des chaînes.

J'ai estimé qu'une paire de Watson-Crick (surtout une paire A.U) n'avait de sens que dans un environnement hélicoïdal. J'ai donc cherché à réaliser la paire A.U dans une configuration qui n'implique pas de torsion dans la chaîne des phosphates, et un environnement empilé. Ceci implique deux conditions :

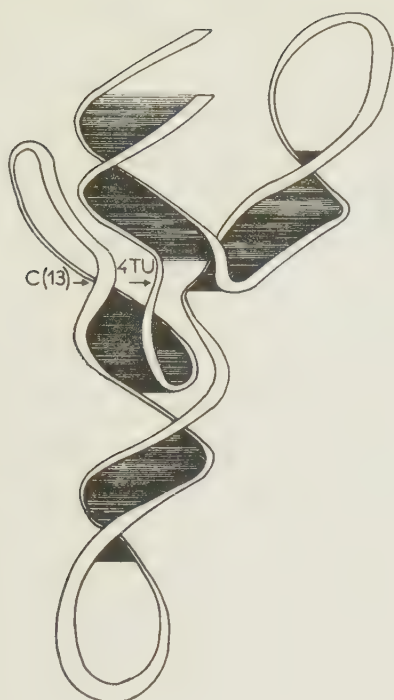


Fig. 7-a.
Le premier modèle
4-2-2 .

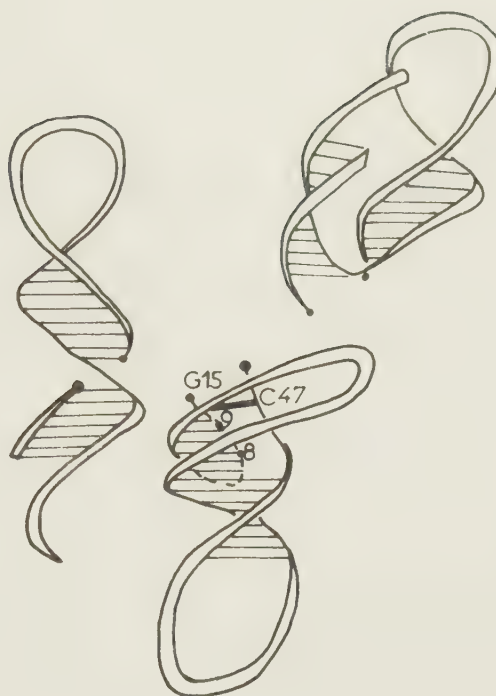


Fig. 7-b.
La proximité des résidus
15 et 47 est compatible,
soit avec un modèle 2-4-2,
soit avec un modèle 4-2-2 .

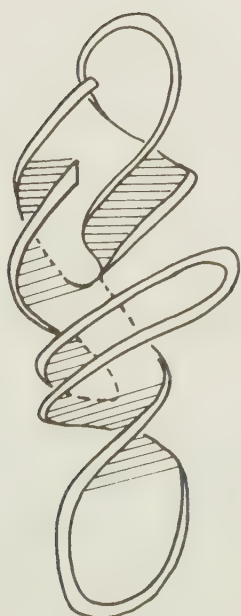


Fig. 7-c.
Le second modèle 4-2-2 .

(dans le premier modèle, il enveloppait le bras DiHU)

- la purine de la paire G.C ou A.U est retournée à 180° par rapport à sa position ordinaire.

Une fois définie cette région autour du bras DiHU, deux solutions distinctes apparaissaient clairement :

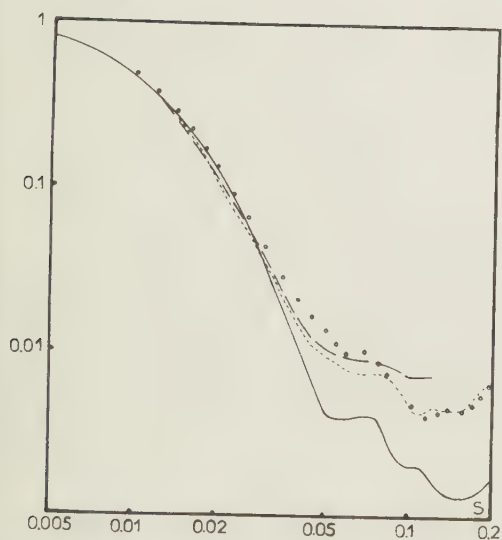
- a) un modèle 2-4-2, mais différent de tous ceux proposés par ailleurs, l'extrémité acceptrice venant près de l'anticodon et la boucle GTTC se plaçant à l'extrémité opposée de la molécule.
- b) un modèle 4-2-2 présentant la particularité nouvelle suivante : la boucle GTTC est très proche de l'extrémité acceptrice et peut éventuellement l'envelopper.

La deuxième solution permet évidemment un Rg plus grand. Elle présente le second avantage de fournir une explication à l'inaccessibilité des nucléotides de la boucle GTTC. Cette inaccessibilité n'était pas dans les prémisses mais au contraire s'est déduite d'hypothèses différentes.

Si effectivement, la boucle GTTC enveloppe le bras accepteur, le tARN est dans une situation qui confirme les conceptions de Danchin sur la présence d'un noeud élastique.

Remarquons aussi qu'il est pas exclu que le blocage de la boucle puisse s'expliquer par des liaisons hydrogène supplémentaires et plus précisément par des interactions entre cette boucle et la boucle DiHU. Cependant, même si cela est vrai, l'étude que nous avons menée montre que, pour l'instant, on peut se passer d'une telle explication.

Enfin, le modèle 4-2-2 a été construit en tenant compte d'un nombre restreint de données, mais considérées comme sûres et importantes. On peut signaler que la plupart des modèles proposés par ailleurs tiennent compte de la réactivité chimique de telle ou telle base, mais sont erronés du point de vue de la diffusion centrale des rayons X (64,124,125,126) ou du point de vue de la réalisabilité de la liaison thioU-C(13,44).



Courbes théoriques de
modèles 4-2-2.

Fig. 7-d.

o : modèle préliminaire,
compact dans sa partie
supérieure, mais où le doigt
ne fait pas corps avec le
reste de la molécule.
---premier modèle 4-2-2.
les deux autres courbes
sont les courbes expéri-
mentales pour le tARN Val
chargé et non chargé.

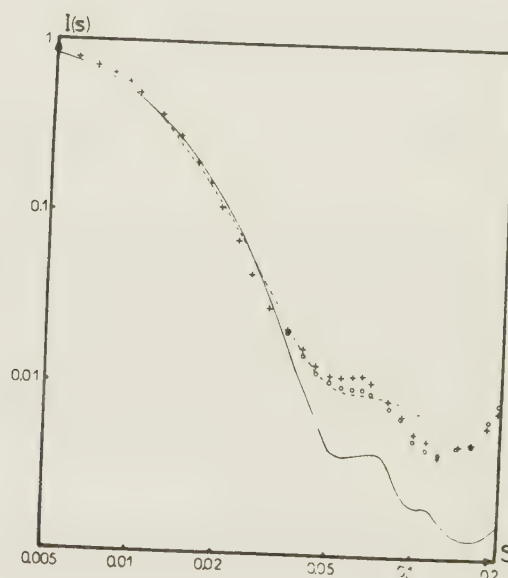


Fig. 7-e.

+ deuxième modèle 4-2-2, avec
les coordonnées de l'ARN IO .
o deuxième modèle 4-2-2 avec
les coordonnées de l'ARN II .

CHAPITRE 8

LA RECONNAISSANCE CODON-ANTICODON.

Un des espoirs placés dans la structure du tARN est qu'elle pourrait jeter quelques lumières sur l'origine et la nature du code génétique. Selon Crick, le catalogue qui fait correspondre les codons aux acides aminés est dans une large mesure arbitraire, tandis que les propriétés générales de dégénérescence du code reflètent une propriété de l'association codon-anticodon. Dans l'hypothèse du wobble, Crick propose que les deux premières paires de bases de l'association se forment de manière rigide, tandis qu'une certaine flexibilité existerait en troisième position. Cette flexibilité pourrait être attribuée à la structure particulière de la boucle de l'anticodon, comme cela a été proposé dans le modèle mécaniste de Fuller et Hodgson.

Cependant, il existe au moins deux catégories générales de faits qui ne cadrent pas entièrement avec l'hypothèse du wobble.

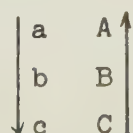
1) Dans les régions ordonnées de l'arrangement en trèfle, les paires non complémentaires G.U ou même C.A sont fréquemment incorporées. Cette observation suggère qu'une flexibilité des bases dans une double-hélice d'ARN est plutôt la règle que l'exception, et que notre attention devrait plutôt se porter sur les raisons qui permettent à la reconnaissance d'être stricte sur les deux premières positions.

2) La présence d'inosine dans certains anticodons de tARNs de levure ne peut être expliquée par des raisons d'économie, puisqu'on connaît des cas d'espèces multiples de tARNs possédant le même anticodon et que la redondance du génôme dans la levure qui correspond au tARN est considérable (127).

J'ai proposé une manière nouvelle d'aborder le problème de la reconnaissance codon-anticodon qui s'appuie sur des considérations énergétiques.

ENERGIE DE L'INTERACTION CODON-ANTICODON.

Considérons deux triplets en interaction :

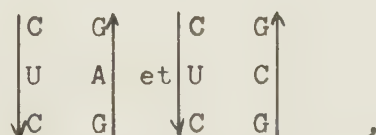


Le codon est écrit à gauche, l'anticodon à droite, et les flèches vont de l'extrémité 5' à l'extrémité 3'. La troisième position de l'association (cC) contient la troisième base du codon et la première de l'anticodon. Le wobble considère uniquement les interactions horizontales, "dans le plan" : a-A, b-B et c-C. On ne peut négliger complètement les interactions verticales (a-b, b-c, etc.) car la configuration du codon ou de l'anticodon au moment de l'approche en dépend. La capacité d'un anticodon à fonctionner dans le processus de traduction est en partie conditionnée par ces interactions, et de manière générale, par la structure de toute la boucle.

Ici, nous nous intéressons surtout à une autre catégorie d'interactions, celle qui résulte du recouvrement partiel d'une base du codon avec une base de l'anticodon qui ne lui est pas associée : ce sont les interactions diagonales a-B, b-C, etc. Celles-ci peuvent contribuer à l'énergie libre dégagée lors de la formation de l'association codon-anticodon.

Dans cette perspective énergétique, la présence ou l'absence d'ambiguïtés de codage est liée à la possibilité de séparer suffisamment un certain nombre de niveaux d'énergie qui dépend du nombre de codons en présence, du nombre d'anticodons et de la dégénérescence du code.

Connaissant le code génétique, on peut se demander si entre les deux associations ci-dessous,



la différence d'énergie est telle que le codon CUC ne s'associera pas à l'anticodon GCG dans plus de, disons, 1 % des cas ?

Si u est l'énergie libre de l'association, et si la proportion des diverses associations en présence, relatives à un même codon, suit une loi en $\exp(-u/kT)$, une différence de 4 kcal. entre deux niveaux correspondra à 0,15 % de traduction erronée à 37°C, valeur élevée au regard de la précision du processus de biosynthèse des protéines.

Des calculs, par la mécanique quantique, d'énergies d'interaction entre triplets dans une configuration donnée conduisent à l'idée que la contribution énergétique "dans le plan" de formation d'une paire A.U peut être du même ordre de grandeur que les contributions diagonales⁽¹²⁸⁾. Plus précisément, dans la configuration de l'ARN II, Pierre Claverie évalue à -5 kcal. l'énergie de la paire A.U et fournit les valeurs suivantes pour les interactions diagonales de guanines :



Ceci nous suggère que la reconnaissance de triplets dans la configuration considérée est nécessairement ambiguë.

LA RECONNAISSANCE RIGIDE EST AMBIGUE.

Si, dans une paire A.U, on substitue C à A sans déplacer la chaîne des sucres-phosphates, aucun contact court n'en résulte entre les deux bases (Fig. 8-c). Les calculs de P. Claverie indiquent que dans cette configuration, la paire C.U est légèrement attractive. Compte-tenu des interactions diagonales, il en résulte que, dans une vingtaine de cas, l'association d'un triplet donné avec un triplet qui ne lui est pas complémentaire se situe à moins de 4 kcal. de l'association correcte. Par suite des limitations inhérentes à ce genre de calculs, les conséquences qu'on en tire sont très qualitatives. Cependant, elles suggèrent qu'une traduction basée sur des associations de Watson-Crick dans une configuration rigide avec empilement de bases a de fortes chances d'être ambiguë.

L'ARN EN DOUBLE-HELICE N'EST PAS RIGIDE.

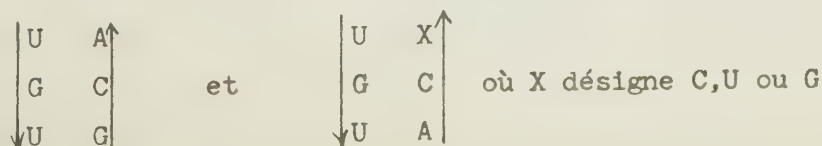
En effet, plus de 5% des paires de bases dans les segments ordonnés du tARN sont non-complémentaires (tableau 8-b). Un examen attentif des séquences suggère que ces appariements non-complémentaires sont présents en vertu des lois générales d'association des oligonucléotides, qu'ils ne doivent pas être considérés comme des "défauts", qui seraient requis par exemple pour la reconnaissance par les synthétases. C'est ainsi que la première association du tableau 8-b est rencontrée, non seulement dans 6 tARNs aux endroits les plus variés, mais également dans la "queue" de l'ARN 5s ribosomal humain⁽¹⁴⁵⁾.

LA RECONNAISSANCE NON-RIGIDE EST TRES AMBIGUE.

Considérons maintenant la reconnaissance telle qu'elle est décrite par le wobble : rigide dans les deux premières positions, et flexible dans la troisième. Alors, il est facile de constater que les ambiguïtés prédites par le modèle de reconnaissance rigide sur les trois positions seront intégralement maintenues. Si, par exemple, la différence d'énergie entre les deux associations



est faible, alors cela est encore plus vrai de la différence entre



	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Leu	Ser	Term	Trp	C
					A
C	Leu	Pro	His	Arg	U
			Gln	Arg	C
					A
A	Ile	Thr	Asn	Ser	U
	Met		Lys	Arg	C
					A
G	Val	Ala	Asp	Gly	U
			Glu		C
					A

Fig. 8-a .Diagrammes de reconnaissance de codons par des tARNs de provenances variées. Ref. 139- 135
Les cercles sont relatifs à E.coli, les triangles à la levure et les carrés au cochon d'Inde.
Les codons "reconnus" par un même tARN sont joints par un trait vertical.

association	espèce de tARN	provenance	ref.
	tRNA ^{Val}	(Baker's yeast)	136
		(<i>Torulopsis utilis</i>)	137
	tRNA ^{Ala}	(<i>E. coli</i>)	138
	tRNA ^{Ser}	(Baker's yeast)	10
	tRNA ^{Met}	(Brewer's yeast)	11
		(<i>E. coli</i>)	139
	tRNA ^{Ala}	(Baker's yeast)	10
	tRNA ^{Val}	(Baker's yeast)	136
		(<i>Torulopsis utilis</i>)	137
	tRNA ^{Phe}	(Baker's yeast)	140
	tRNA ^{Tyr}	(Baker's yeast)	12
	tRNA ^{Met}	(<i>E. coli</i>)	141
	tRNA ^{Phe}	(wheat germ)	142
	tRNA ^{Phe}	(<i>E. coli</i>)	143
	tRNA ^{Ile}	(<i>Torulopsis utilis</i>)	144

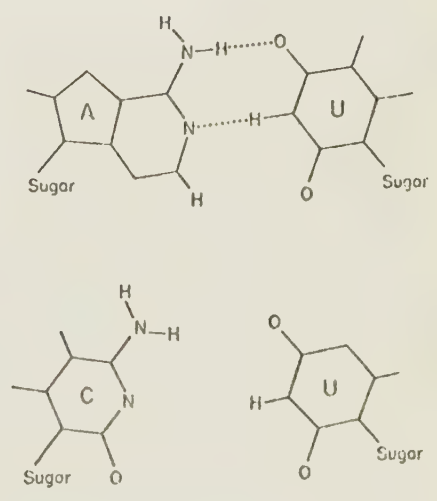


Fig. 8-c .Les paires A.U et C.U dans la même position "rigide".

Tableau 8-b. Les associations non complémentaires dans les tARNs.

Ainsi, le codon UGU de la cystéine reconnaîtrait en même temps l'anticodon GCA prévu par le wobble, et un anticodon possible pour l'arginine : ACG . Mais ceci n'est pas le cas *in vivo* . On peut alors envisager deux solutions :

1) Notre ensemble d'hypothèses est incorrect : il existerait des enzymes de vérification pour l'association, ou bien, celle-ci se produit dans une configuration peu empilée, qui permet une reconnaissance séparée dans chaque position.

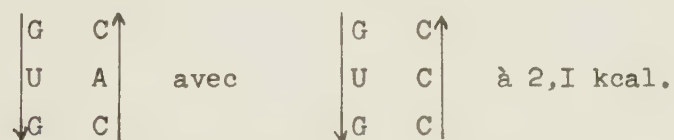
2) Le point de vue est essentiellement correct, et l'anticodon ACG n'existe pas. Une mutation qui produirait un tARN muni d'un tel anticodon en quantités appréciables serait létale.

EXAMEN DES AMBIGUITES POSSIBLES ET DE LEURS CONSEQUENCES.

La configuration réelle de l'association codon-anticodon n'est pas connue. Par conséquent, il est difficile de prédire exactement les ambiguïtés susceptibles de se produire. On peut cependant chercher des suggestions par différents côtés.

A) MECANIQUE QUANTIQUE.

Une erreur suggérée par les calculs de Claverie est la suivante :



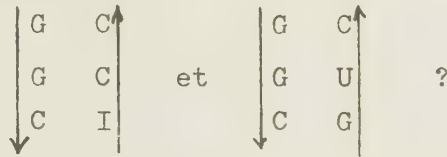
On en déduit que l'anticodon CCC n'existe pas. Il n'a pas été observé de tARN^{Gly} se comportant comme un accepteur avec l'anticodon CCC dans le foie de cochon d'Inde (Fig. 8-a) . De même, dans la levure, trois fractions de tARN^{Gly} ont été décrites, dont aucune ne possède cette propriété. Cependant, les auteurs interprètent leurs résultats en disant que leur fraction III est un mélange contenant un tel accepteur. On voit par cet exemple que certains des résultats qui ont pu être apportés à l'appui du wobble ,adoptaient en réalité le wobble comme hypothèse implicite. Une étude récente faite sur les fractions tARN^{Gly} de E. coli montrent la présence d'une fraction qui se comporte exactement comme la fraction III de levure. Il n'est pas dit par conséquent que l'interprétation des résultats par l'hétérogénéité de la fraction III soit correcte.

B) STRUCTURE DU tARN.

A partir de l'association remarquable entre GUC et CGG, on peut essayer de construire une ambiguïté:



Mais comme une paire G.U a été remplacée par une paire G.C beaucoup plus forte, la possibilité d'ambiguïté est peu préoccupante. Qu'advient-il cependant si on substitue de l'inosine à la guanine dans la première association ? Quelle est la séparation des niveaux d'énergie entre



On retrouve alors l'observation remarquable que dans la levure ou le cochon d'Inde, il n'existe pas d'anticodon ICC alors que rien ne l'interdit dans l'hypothèse du wobble.

C) BIOSYNTHESE DES PROTEINES.

On pourrait citer le cas de l'ambiguïté leucine-phénylalanine dans les systèmes de biosynthèse acellulaires qui accrédite l'idée de reconnaissance non rigide sur la première position de l'association. Un cas plus intéressant est suggéré par le fait que le $\text{tARN}^{\text{fMet}}$ qui se fixe en un site particulier sur le ribosome répond au codon GUG de la valine, alors que son anticodon est CAU. Une première interprétation, compatible avec le wobble, serait qu'il s'agit là d'un tARN possédant une structure tout à fait particulière, ou que la reconnaissance sur le premier site ribosomal est différente de celle sur le suivant.

Dans le cadre des idées présentées ici, on considère au contraire que la flexibilité dans la première position de l'association est chose naturelle. Comment se fait-il alors que le tARN^{Met} ne se fixe pas sur le codon GUG de la valine ? La réponse que nous suggérons ici coule de source : le tARN^{Met} possède un anticodon C^+AU dans lequel la première base est modifiée. La prédiction que nous faisons est alors que cette modification va dans le sens d'un affaiblissement de la liaison de la paire G.C.

En examinant tous ces cas, l'un après l'autre, on se rend compte que la reconnaissance codon-anticodon est une suite de cas particuliers, et que la belle simplicité prédite par l'hypothèse du wobble ne semble pas réalisée dans la cellule. Notre hypothèse semble mieux rendre compte de la complexité des phénomènes observés. En particulier, elle établit une distinction entre les paires G.U et les paires U.G (où la première lettre désigne la base sur le codon). En effet, G.U est un substitut à G.C, tandis que U.G est un substitut à U.A. Mais comme la paire G.C est beaucoup plus énergétique que la paire U.A, l'utilisation des paires G.U en troisième position est plus dangereuse que celle des paires U.G : elle peut exclure la possibilité d'utiliser quelque part ailleurs, un anticodon qui commence par C. Une dissymétrie dans l'utilisation des paires U.G et G.U

est effectivement observée dans le code. Elle peut être liée à cette remarque.

L'HYPOTHESE DU TRIPLET MANQUANT.

En résumé, on peut proposer les formules suivantes :

- 1) La reconnaissance rigide de triplets dans une configuration empilée est ambiguë.
- 2) La reconnaissance non rigide de triplets dans une configuration voisine de celle des segments ordonnés du tARN est ambiguë.
- 3) Dans la configuration réelle de l'association codon-anticodon, le code serait très ambigu si les 61 anticodons naturels étaient présents sur leurs tARNs correspondants.
- 4) Le code n'est pas ambigu parce que de nombreux anticodons possibles n'existent pas et que d'autres sont chimiquement modifiés.

Les implications de cette hypothèse concernant l'origine du code génétique sont importantes, elles ne seront pas discutées ici. Par ailleurs, elle soulève à nouveau une des questions centrales dans le problème des origines de la vie : comment peut-on concevoir l'évolution d'un système à réplication fidèle vers un système à réplication rigoureuse ?

Enfin, signalons que les idées avancées suggèrent de nombreuses expériences⁽⁴⁹⁾ dont certaines sont à la portée des techniques biochimiques actuelles.

CHAPITRE 9.

CONCLUSION

Une des ambitions de la biologie moléculaire est d'expliquer le fonctionnement de la cellule par la structure des macromolécules impliquées dans les processus variés, de régulation, de transfert d'information, de catalyse, de perméation, etc. L'ARN de transfert occupe une place essentielle dans la traduction de l'information génétique ; il était même défini fonctionnellement avant même d'avoir été identifié. Pour l'essentiel, on lui attribue une double fonction de reconnaissance :

1) - reconnaissance (directe, ou indirecte) de l'acide aminé. Dans les cellules actuelles, il est à peu près certain que la reconnaissance est essentiellement indirecte, le tARN et l'acide aminé étant fixés séparément par l'enzyme d'activation.

2) - reconnaissance (directe ou indirecte) du codon. Quoique les conditions générales de cette reconnaissance (sur le ribosome, dans un certain environnement) ne dépendent pas que des réactants, ceci étant vrai de toute réaction cellulaire, les hypothèses de reconnaissance directe de l'ARN de transfert par le codon sont jusqu'à présent suffisantes.

La dissociation entre ces deux fonctions a été démontrée par les expériences déjà classiques de Chapeville et al.⁽¹⁴⁶⁾. D'autres fonctions ont pu être invoquées, dans la régulation des chaînes métaboliques de biosynthèse des acides aminés, dans la régulation de la vitesse de synthèse des protéines, dans la différenciation cellulaire. Cependant, les fonctions essentielles sont les deux qui ont été signalées au début.

La partie expérimentale du travail présenté s'est attachée exclusivement au volet structural de la question. Les premières expériences de diffusion centrale des rayons X ont contribué à dégrossir l'image. Le rayon de gyration, le volume, l'allure de l'intensité diffusée conduisaient à un objet d'allure cylindrique, de 80 Å de long pour 25 Å de diamètre, donc plus compact que l'épingle à cheveux, mais cadrant mal avec la représentation planaire du trèfle. Par ailleurs, si les biochimistes, utilisant des techniques sensibles, estiment que la conformation "native", active de l'ARN de transfert n'existe que dans une gamme étroite de conditions expérimentales, l'expérience accumulée dans les mesures de rayons X par contre suggèrent que le polymorphisme dans le tARN affecte le plus souvent des aspects subtils de la structure secondaire plutôt que la disposition spatiale des quatre bras de la molécule. Aucune des expériences de rayons X

effectuées dans ce laboratoire n'a conduit par exemple à une courbe qui présenterait une épaule vers $s=0,03$ caractéristique d'une classe assez large de repliements du trèfle. La seule différence importante observée se trouve dans les queues de courbes, entre le tARN chargé et le tARN non chargé ; son interprétation est délicate, et il n'est pas exclu que l'on touche là aux limites de la technique de diffusion centrale qui, comme son nom l'indique, s'est surtout consacrée à l'interprétation des signaux aux petits angles.

On pouvait ne pas s'estimer satisfaits de la détermination de quelques paramètres physiques, et de leur comparaison. Savoir que l'ARN de transfert est en cylindre plutôt qu'en boule, en quoi cela permet-il de mieux le comprendre ? Evidemment, ces déterminations de paramètres pouvaient renvoyer à d'autres expériences physiques comme les équilibres de sédimentation, ou les dénaturations thermiques, mais n'apportaient pas encore d'éclaircissement sur la fonction du tARN.

Plus grave encore, il existait à première vue une contradiction entre cette forme en cylindre et les données chimiques qui suggéraient un arrangement en trèfle de la molécule. Cette contradiction s'est avérée difficile à résoudre. Plusieurs équipes utilisant les mêmes techniques physiques que nous, ont préféré en quelque sorte ne pas faire confiance à leurs données. Car si toutes les mesures expérimentales de rayon de gyration faites sur des accepteurs purifiés chargé ou non chargé donnent une valeur supérieure à 24 Å (le présent travail est pour l'instant le seul qui ait considéré, par diffusion des rayons X, la forme chargée) - par contre, les modèles présentés par les autres groupes ont tous un rayon de gyration inférieur à 23 Å. Il convient de signaler qu'en matière de rayon de gyration, un gain d'un angström n'est pas toujours aisé à obtenir. Si, dans les régions ordonnées, on écarte les plateaux de bases de 3,4 au lieu de 3,0 Å, il en résulte un gain de 0,7 Å dans le rayon de gyration, et cette valeur est également la marge dont on dispose en changeant les boucles.

Il est vrai que la construction de modèles n'est pas une science codifiée mais une technique en élaboration qui requiert sans doute de la rigueur, mais aussi de l'imagination. Et j'avoue sans fausse honte que je n'aurais jamais imaginé l'arrangement 4-2-2 si une donnée chimique particulière, celle de l'interaction thio U - C n'était venue imposer une contrainte qui se trouvait tout à fait en dehors du domaine envisagé.

Ainsi, il a été possible de concevoir un modèle où trois des régions hélicoïdales du tARN sont empilées les unes sur les autres. Ce modèle a été présenté pour la première fois à une réunion de l'EMBO sur l'ARN de transfert, en mars 1969. Les arguments chimiques apportés par Yaniv, Favre et Barrell joints aux arguments de rayons X ont largement

contribué à chasser des esprits les modèles 4-4 et 6-2, et à restreindre le champ aux modèles dans lesquels trois régions hélicoïdales sont empilées.

On arrivait ainsi à une image assez précise. Quelle conclusion pouvait-on en tirer ?

D'abord, on pouvait affirmer de manière à peu près certaine que l'anticodon et le bras accepteur se trouvent aux deux extrémités de la molécule ; que l'anticodon n'interagit avec aucune autre partie de la molécule, mais que pour l'extrémité acceptrice, la conclusion était plus réservée. Par conséquent, s'il y a jamais eu une forme de reconnaissance directe de l'acide aminé par la région de l'anticodon, celle-ci s'est perdue au cours de l'évolution. D'autre part, on ne peut pas exclure une reconnaissance primitive de l'acide aminé par une région autre que celle de l'anticodon, et dans ce cas, le modèle 4-2-2 suggérerait pour candidat la boucle GTTC qui, à l'heure actuelle semble être une des régions du tARN qui varie le moins, la moins capable, par sa séquence, de souplesse cognitive, et à laquelle on attribue volontiers un rôle dans la fixation du tARN par le ribosome.

On voit par conséquent que le lien à établir entre structure et fonction, même pour le tARN n'est pas simple, d'autant plus qu'une des fonctions définissant le tARN est maintenant en quelque sorte transférée aux enzymes d'activation.

On voit également qu'il n'est pas possible, pour une question aussi fondamentale de se contenter d'une image figée des rapports entre les constituants cellulaires, et que l'explication scientifique doit inclure le processus évolutif qui a donné naissance à ces relations.

C'est pourquoi, il est intéressant, au-delà de la détermination précise d'une structure moléculaire, de comprendre les acides nucléiques de manière un peu plus générale, et s'interroger sur leurs potentialités a priori. Ainsi se justifie entre autre l'étude menée sur les propriétés des représentations d'acides nucléiques. Un de ses résultats est de suggérer la possibilité d'existence, dans le passé, d'une forme de tARN à trois filaments. Cependant, la transition d'une telle forme au modèle 4-2-2 n'est pas évidente. Un fait nouveau, apporté par l'étude la plus récente de constructions de modèles (Ch. 7) autorise peut-être à faire le pont. En effet, lors de la construction du dernier modèle 4-2-2, un fait non prévu est apparu de manière frappante : la possibilité d'existence de deux formes distinctes répondant aux mêmes contraintes de départ, l'une étant la forme 4-2-2, et l'autre une forme 2-4-2 inversée par rapport à celle de Levitt, dans laquelle l'extrémité acceptrice se rapproche de l'anticodon. Cependant,

de telles idées ne peuvent être avancées à la légère, notamment parce que toute évolution dans l'appareil traducteur obéit à une logique évolutive propre, qu'il faut expliciter. Mais ceci sort du cadre de la thèse.

En ce qui concerne l'autre fonction du tARN, celle d'être reconnu par le codon, on pourrait penser que le lien avec la structure serait facile à établir. Une fois que les règles d'appariements des bases dans l'association codon-anticodon eurent été énoncées par Crick, il a été généralement admis que ces règles constituaient l'explication même de la structure du catalogue codons-acides aminés, plutôt qu'une conséquence de la structure du code.

Cette attitude était même précisée par un modèle purement mécaniste de la reconnaissance codon-anticodon, dans lequel l'explication était fournie d'emblée, dans la structure de la boucle de l'anticodon du tARN.

Or, les idées que j'ai avancées dans ce domaine vont à l'encontre de la possibilité d'expliquer directement la fonction par la structure. Elles montrent par exemple que la reconnaissance ne doit pas être comprise comme une propriété d'association de deux molécules isolées, mais que le tout réagit sur la partie d'une manière dont on doit et peut tenir compte dans les raisonnements. De plus, la structure physique générale de l'anticodon ne suffit pas, et il faut tenir compte de tous les correctifs chimiques qui lui sont apportés, eux-mêmes fruits de l'évolution, tout en étant dans une large mesure conditionnées par les lois physiques générales d'association des segments d'acides nucléiques.

Par conséquent, le travail présenté ici, en même temps qu'il contribue à préciser les détails du fonctionnement cellulaire, montre que le lien entre structure et fonction n'est pas aussi simple qu'il pouvait paraître et de là, sans prétendre y réussir, suggère que l'explication scientifique se situe au-delà de la simple interprétation mécaniste des phénomènes. Ce travail n'est pas clos, il en appelle d'autres.

CHAPITRE 10

BIBLIOGRAPHIE.

- 1 Hoagland, M.B., Zamecnik, P.C. & Stephenson, M.L. (1957), *Biochim. Biophys. Acta*, 24, 215
- 2 Crick, F.H.C. (1966), *Cold Spr. Harb. Symp. Quant. Biol.*, 31, 3
- 3 Tissières, A. (1959) *J. Mol. Biol.* 1, 365
- 4 Spencer, M., Fuller, W., Wilkins, M.H.F., & Brown, G.L. (1962) *Nature* 194, 1014
- 5 McCully, K.S. & Cantoni, G.L. (1962) *J. Mol. Biol.* 5, 497
- 6 Fresco, J.R. (1963) *Informational Macromolecules*, Ed. H.J. Vogel, Bryson & Lampen
Acad. Press, New-York, p. 121
- 7 Witz, J. (1964) Thèse, Université de Strasbourg.
- 8 Fresco, J.R., Alberts, B.M. & Doty, P. (1960) *Nature*, 188, 98
- 9 Guschlbauer, W. (1966) *Nature*, 209, 258
- 10 Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H.,
Penswick, J.R. & Zamir, A. (1965) *Science* 147, 1462
- 11 Zachau, H.G., Dütting, G. & Feldmann, H. (1966) *Angew. Chemie* 78, 392
- 12 Madison, J.T., Everett, G.A. & Kung, H. (1966) *Science*, 153, 531
- 13 Lake, J.A. & Beeman, W.W. (1968) *J. Mol. Biol.* 31, 115
- 14 Lindahl, T., Adams, A. & Fresco, J.R. (1966) *Proc. Natl. Acad. Sci., Wash.*, 55, 941
- 15 Gartland, W.J. & Sueoka, N. (1966) *Proc. Natl. Acad. Sci., Wash.*, 55, 948
- 16 Stern, R., Zutra, L.E. & Littauer, U.Z. (1969), *Biochemistry*, 8, 313
- 17 Kaji, H. & Tanaka, Y. (1967) *Biochim. Biophys. Acta*, 138, 642
- 18 Gantt, R.R., Englander, S.W. & Simpson, M.V. (1969) *Biochemistry*, 8, 475
- 19 Sarin, P.S. & Zamecnik, P.C. (1965) *Biochim. Biophys. Res. Com.*, 20, 400
- 20 Cohn, M., Danchin, A. & Grunberg-Manago, M. (1969) *J. Mol. Biol.*, 39, 199
- 21 Goodman, H.M., Abelson, J., Landy, A., Brenner, S. & Smith, J.D. (1968), *Nature*, 217, 1019
- 22 Stent, G.S. & Brenner, S. (1962), *Proc. Natl. Acad. Sci., Wash.*, 42, 2005
- 23 Ames, B.N. & Hartman, P.E. (1963), *Cold Spr. Harb. Symp. Quant. Biol.*, 28, 349
- 24 Ibuki, F., Gasior, G. & Moldave, K. (1966), *J. Biol. Chem.*, 241, 2188
- 25 Crick, F.H.C. (1966), *J. Mol. Biol.*, 19, 548
- 26 Fuller, W. & Hodgson, A. (1968), *Nature*, 215, 817

- 7 Crick, F.H.C. (1969), *J. Mol. Biol.* , 38, 367
- 8 Clark, B.F.C., Doctor, B.P., Holmes, K.C., Klug, A., Marcker, K.A., Morris, S.J. & Paradies,
H.H. (1968) , *Nature*, 219, 1222
- 9 Hampel, A., Labanauskas, M., Connors, P.G., Kirkegaard, L., RajBhandary, U.L., Sigler, P.B.,
& Bock, R.M. (1968), *Science*, 162, 1384
- 0 Cramer, F., v.d. Haar, F., Saenger, W., & Schlimme, E. (1968) , *Angew. Chem. Int. Ed.*
Engl. 7, 895
- 1 Kim, S.H. & Rich, A. (1968) *Science*, 162, 1381
- 2 Fresco, J.R., Blake, R.D. & Langridge, R. (1968), *Nature*, 220 , 1285
- 3 Vold, B.S. (1969) , *Biochim. Biophys. Res. Com.* , 35, 222
- 4 Igo-Kemenes, T. & Zachau, H.G. (1971) , *Eur. J. Biochem.* , 18, 292
- 5 Thiebe, R. & Zachau, H.G. (1970) , *Biochim. Biophys. Acta*, 217, 294
- 6 Litt, M. (1969), *Biochemistry*, 8, 3249
- 7 Chuguev, I.I., Axel'rod, V.D. & Baev, A.A. (1970), *Biochim. Biophys. Res. Com.*, 41, 108
- 8 Schmidt, J., Buchardt, B. & Reid, B.R. (1970), *J. Biol. Chem.* 245, 5743
- 9 Seno, T., Kobayashi, M. & Nishimura, S. (1969) , *Biochim. Biophys. Acta*, 174, 408
- 0 Mirzabekov, A.D., Kazarinova, L.Y., Lastity, D. & Baev, A.A. (1969), *FEBS Lett.*, 3, 268
- 1 Philippsen, P., Thiebe, R., Wintermeyer, W. & Zachau, H.G. (1968) , *Biochem. Biophys.*
Res. Com. , 33, 922
- 2 Ninio, J., Favre, A. & Yaniv, M. (1969), *Nature*, 223, 1333
- 3 Levitt, M. (1969) , *Nature*, 224, 759
- 4 Connors, P.G., Labanauskas, M & Beeman, W.W. (1969), *Science*, 166, 1528
- 5 Danchin, A. (1971), *FEBS Lett.*, 13, 152
- 6 Fuller, W., Arnott, S. & Creek, J. (1969), *Biochem. J.*, 114 , 26p
- 7 Koltun, W.L. (1965), *Biopolymers*, 3, 665
- 8 Yaniv, M., Favre, A. & Barrell, B.G. (1969) , *Nature*, 223, 1331
- 9 Ninio, J. (1971) , *J. Mol. Biol.*, 56, 63
- 0 Luzzati, V., Witz, J. & Baro, R. (1963) *Le Journal de Physique*, 24, I41A
- 1 Luzzati, V. (1957), *Acta Cryst.* , 10, 136
- 2 Luzzati, V. (1958), *Acta Cryst.*
- 3 Feigin, L.A. (1967) "Soviet Physics"-Cristallography, 12, 228
- 4 Hossfeld, F. (1968) , *Acta Cryst.*, A24, 643
- 5 Lake, J.A. (1967), *Acta Cryst.*, 23, 191
- 6 Langridge, R., Marvin, D.A., Seeds, W.E., Wilson, H.R., Hooper, C.W., Wilkins, M.H.F.,
& Hamilton, L.D. (1960) , *J. Mol. Biol.* , 2, 38
- 7 Mittelbach, P. & Porod, G. (1961) , *Acta Physica Austriaca* , 2, 185
- 8 Kirste, R.G. & Oberthür, R.C. (1969), *Die Makromolekulare Chemie* , 127, 301
- 9 Bram, S. & Beeman, W. (1971) *J. Mol. Biol.* , 55, 311
- 0 Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. & Hill, R.L.A. (1969)
J. Mol. Biol. , 42, 65

- 1 Welton, M.G.E. & Pelc, S.R. (1966), *Nature*, 209, 870
- 2 Lacey, J.A. & Pruitt, K.M. (1969), *Nature*, 223, 799
- 3 Cantor, C.R., Jaskunas, S.R. & Tinoco, I. (1966), *J. Mol. Biol.*, 20, 39
- 4 Cramer, F., Doepner, H., v.d. Haar, F., Schlimme, E. & Seidel, H. (1968), *Proc. Natl. Acad. Sci, Wash.*, 61, 1384
- 5 Conçus par Ivor Smith.
- 6 Fuller, W., Hutchinson, F., Spencer, M. & Wilkins, M.H.F. (1967), *J. Mol. Biol.*, 3, 507
- 7 Arnott, S., Wilkins, M.H.F., Fuller, W. & Langridge, R. (1967), *J. Mol. Biol.*, 3, 535
- 8 Arnott, S., Fuller, W., Hodgson, A. & Prutton, I. (1968), *Nature*, 220, 561
- 9 Arnott, S., Dover, S.D., & Wonacott, A.J. (1969), *Acta Cryst.*, B25, 2192
- 0 Milman, G., Chamberlain, M & Langridge, R. (1967), *Proc. Natl. Acad. Sci, Wash.*, 57, 1084
- 1 Fuller, W., Wilkins, M.H.F., Wilson, H.R., Hamilton, L.D., & Arnott, S. (1965), *J. Mol. Biol.*, 12, 60
- 2 Arnott, S. & Hukins, D.W.L. (1969), *Nature*, 224, 886
- 3 Arnott, S. (1970) *Progress in Biophysics and Molecular Biology*, 21, 267
- 4 Doctor, B.P., Fuller, W. & Webb, N.L. (1969), *Nature*, 221, 58
- 5 Crick, F.H.C. (1967), *Nature*, 213, 798
- 5b Ninio, J. (1971), *Biochimie, sous presse.*
- 6 Cantor, C.R. (1967), *Nature*, 216, 513
- 7 Raacke, I.D. (1968), *Biochim. Biophys. Res. Com.*, 31, 528
- 8 Madison, J.T. (1968), *Ann. Rev. Biochem.*, 37, 131
- 9 Lewis, J.B. & Doty, P. (1970), *Nature*, 225, 5110
- 0 Spirin, A.S. (1960), *J. Mol. Biol.*, 2, 436
- 1 Fresco, J.R., Alberts, B.M. & Doty, P. (1960), *Nature*, 188, 98
- 2 Thomas, C.A. & Mac Hattie, L.A. (1967), *Ann. Rev. Biochem.*, 36, 485
- 3 Cairns, J. (1963), *Cold. Spr. Harb. Symp. Quant. Biol.*, 28, 43
- 4 Ris, H. & Chandler, B.L. (1963), *Cold Spr. Harb. Symp. Quant. Biol.*, 28, I
- 5 Kleinschmidt, A.K., Burton, A. & Sinsheimer, R.L. (1963), *Science*, 142, 961
- 6 Vinograd, J. & Lebowitz, J. (1966), *J. Gen. Physiol.*, 49, 103
- 7 Weiss, B. (1970), *Proc. Natl. Acad. Sci, Wash.*, 65, 652
- 8 Kiger, J.A. Jr., & Sinsheimer, R.L. (1969), *J. Mol. Biol.*, 40, 467
- 9 Bourgaux, P., Bourgaux-Ramoisy, D. & Dulbecco, R. (1969), *Proc. Natl. Acad. Sci, Wash.*, 64, 701
- 0 Gilbert, W. & Dressler, D. (1968), *Cold Spr. Harb. Symp. Quant. Biol.*, 33, 473
- 1 Spiegelman, S. & Haruna, I. (1966), *J. Gen. Physiol.*, 49, 263
- 2 Reanney, D.C. & Ralph, R.K. (1968), *J. Theoret. Biol.*, 21, 217

- 93 Delbrück, M. (1962), Proceedings of Symposia in Applied Mathematics, 14, 55
- 94 Thomas, C.A. & Mac Hattie, L.A. (1964), Proc. Natl. Acad. Sci., Wash., 52, 1297
- 95 Danchin, A. & Guéron, M. (1970), J. Chem. Phys., 53, 3599
- 96 Orgel, L.E. (1968), J. Mol. Biol., 38, 381
- 97 Michelson, A.M. (1963). The Chemistry of Nucleosides and Nucleotides, Acad. Press, London
- 98 Hoogsteen, K. (1959), Acta Cryst., 12, 822
- 99 Hoogsteen, K. (1963), Acta Cryst., 16, 907
- 100 Rich, A., Davies, D.R., Crick, F.H.C. & Watson, J.D. (1961), J. Mol. Biol., 3, 71
- 101 Claverie, P. (1970), communication personnelle.
- 102 Haschemeyer, A.E.V. & Rich, A. (1967), J. Mol. Biol., 27, 369
- 103 Scheit, K.H. & Saenger, W. (1969), FEBS Lett., 2, 305
- 104 Saenger, W. & Scheit, K.H. (1970), J. Mol. Biol. 50, 153
- 105 Richardson, J.P. (1966), J. Mol. Biol. , 21, 83
- 106 Gillam, J., Millward, S., Blew, D., v. Tigerstrom, M., Winner, E. & Tener, G.M. (1967) Biochemistry, 6, 3043
- 107 Timasheff, S.N., Witz, J. & Luzzati, V. (1961), Biophys. J., 1, 525
- 108 Luzzati, V., Witz, J. & Timasheff, S.N. (1962), Acides Ribonucléiques et Polyphosphates - Colloque C.N.R.S. , Strasbourg
- 109 Henley, D.D., Lindahl, T. & Fresco, J.R. (1966), Proc. Natl. Acad. Sci, Wash., 55, 191
- 110 Kay, C.M. & Oikawa, W. (1966), Biochemistry, 5, 213
- 111 Cox, R.A. & Kanagalingam, K. (1967), Biochem. J. , 103, 431
- 112 Luzzati, V., Mathis, A., Masson, F. & Witz, J. (1964), J. Mol. Biol. , 10, 28
- 113 Mathis, A. (1966), Thèse, Université de Strasbourg
- 114 Luzzati, V., Witz, J. & Mathis, A. (1967), in Genetic Elements, ed. by D. Shugar , p.41 London : Acad. Press
- 115 Crick, F.H.C. & Watson, J.D. (1954), Proc. Roy. Soc., A223, 80
- 116 Vawra (1970). Résultat communiqué par R. Oberthür.
- 117 Luzzati, V., Masson, F., Mathis, A. & Saludjian, P. (1967), Biopolymers, 5, 491
- 118 Cohen, G. & Eisenberg, H. (1968), Biopolymers, 6, 1077
- 118b Millar, D.B. (1966), Nature, 211, 1088
- 119 Pilz, I., Kratky, O., Cramer, F., v.d. Haar, F. & Schlimme, E. (1970), Eur. J. Biochem., 15, 401
- 120 Gulik, A., Inoué, H. & Luzzati, V. (1970), J. Mol. Biol., 53, 221
- 121 Danchin, A. & Grunberg-Manago, M., (1970), FEBS Lett., 9, 327
- 122 Adler, A.J. & Fasman, G.D. (1970), Biochim. Biophys. Acta, 204, 183
- 123 Melcher, G., Paulin, D. & Guschlbauer, W. (1971), Biochimie, 1, 43

- 124 Melcher, G. (1969), FEBS Lett., 3, 185
- 125 Wu, T.T. (1969), Bull. Math. Biophysics, 31, 395
- 126 Abraham, D.J. (1971), J. Theoret. Biol., 30, 83
- 127 Schweizer, E., Mac Kechnie, C & Halvorson, H.O. (1969), J. Mol. Biol., 40, 261
- 128 Nash, H.A. & Bradley, D.F. (1965) Biopolymers, 3, 261
- 129 Söll, D., Jones, D.S., Ohtsuka, E., Faulkner, R.D., Lohrmann, R., Hayatsu, H., Khorana, H.G.,
Cherayil, J.D., Hampel, A. & Bock, R.M. (1966), J. Mol. Biol., 19, 556
- 130 Caskey, C.T., Beaudet, A. & Nirenberg, M. (1969), J. Mol. Biol., 37, 99
- 131 Kano-Sueoka, T., Nirenberg, M. & Sueoka, N. (1968), J. Mol. Biol., 35, 1
- 132 Kellog, D.A., Doctor, B.P., Loebel, J.E. & Nirenberg, M.W. (1966), Proc. Natl. Acad.
Sci, Wash., 55, 912
- 133 Nirenberg, M., Caskey, T., Marshall, R., Brimacombe, R., Kellog, D., Doctor, B., Hatfield, D.,
Levin, J., Rottman, F., Peska, S., Wilcox, M., & Anderson, F. (1966), Cold
Spr. Harb. Symp. Quant. Biol., 31, II
- 134 Doctor, B.P., Loebel, J.E. & Kellog, D.A. (1966), Cold Spr. Harb. Symp. Quant. Biol.,
31, 543
- 135 Reeves, R.H., Imura, N., Schwam, H., Weiss, G.B., Schulman, L.H. & Chambers, R.W. (1968),
Proc. Natl. Acad. Sci, Wash., 60, 1450
- 136 Baev, A.A., Venkstern, T.V., Mirzabekov, A.D., Krutilina, A.I., Li, L. & Aksel'rod,
V.D. (1967), Molek. Biol., 1, 754
- 137 Takemura, S., Mizutani, T., & Miyazaka, M. (1968), J. Biochem., 63, 277
- 138 Yaniv, M. & Barrell, B.G. (1969), Nature, 222, 278
- 139 Dube, S.K., Marcker, K.A. & Clark, B.F.C. (1968), Nature, 220, 1039
- 140 RajBhandary, U.L., Chang, S.H., Stuart, A., Faulkner, R.D., Hoskinson, R.M. & Khorana,
H.G. (1967), Proc. Natl. Acad. Sci., Wash., 57, 751
- 141 Cory, S., Marcker, K.A., Dube, S.K. & Clark, B.F.C. (1968), Nature, 220, 1039
- 142 Dudock, B.S., Katz, G., Taylor, E.K. & Holley, R.W. (1969), Proc. Natl. Acad. Sci, Wash.,
62, 941
- 143 Barrell, B.G. & Sanger, F. (1969), FEBS Lett., 3, 275
- 144 Takemura, S., Murakami, M., & Miyazaka, M. (1969), J. Biochem., 65, 489
- 145 Forget, B.G. & Weissman, S.M. (1967), Science, 158, 1695
- 146 Chapeville, F., Lipman, G., von Ehrenstein, B., Weisblum, B., Ray, W.Y.J^r, & Benzer, S. (1962)
Proc. Natl. Acad. Sci, Wash., 48, 1086

